● 国家出版基金资助项目

# 华罗庚文集

## 应用数学卷 I

杨德庄／主编

科学出版社
www.sciencep.com

国家出版基金资助项目

# 华罗庚文集

## 应用数学卷 I

杨德庄　主编

(华罗庚应用数学与信息科学研究中心)

科 学 出 版 社

北　京

## 内 容 简 介

本卷介绍著名数学家华罗庚先生在应用数学领域的成就.

本卷分卷 I、卷 II 两卷, 卷 I 主要内容包括近似分析中的数论方法和应用统计中的数论方法, 卷 II 主要内容包括计划经济大范围最优化数学理论、关于经济优化平衡的数学理论、数学普及之初简介、统筹方法平话及补充、优选法平话及补充、优选学等. 从卷 I、卷 II 可以看出华罗庚在中国发展应用数学的开拓性工作分两个层面: 创造性工作层面与普及推广工作层面, 也可以看出他的探索创新之路和他的深邃的导向观点.

本卷适合数学及相关专业大学生、研究生、教师及科研人员阅读参考.

**图书在版编目(CIP)数据**

华罗庚文集: 应用数学卷 I/杨德庄主编. —北京: 科学出版社, 2010
ISBN 978-7-03-027251-5

I. 华… II. 华… III. ①数学-文集 ②应用数学-文集 IV. 01-53

中国版本图书馆 CIP 数据核字 (2010) 第 069910 号

责任编辑: 张　扬 / 责任校对: 陈玉凤
责任印制: 钱玉芬 / 封面设计: 黄华斌

科学出版社 出版
北京东黄城根北街 16 号
邮政编码: 100717
http://www.sciencep.com

中国科学院印刷厂 印刷

科学出版社发行　各地新华书店经销

*

2010 年 5 月第 一 版　开本: B5(720×1000)
2010 年 5 月第一次印刷　印张: 31 1/4
印数: 1—3 000　字数: 607 000

**定价: 98.00 元**

(如有印装质量问题, 我社负责调换)

纪念华罗庚先生诞辰100周年

# 出版说明

“华罗庚是他那个时代的世界级领袖数学家之一, 对于中国近代数学发展作出了重大贡献”. “他的绝大部分工作时间是在中国度过的. 如果今天有许多中国数学家能在科学前沿作出突出贡献的话, 如果数学在中国能享有异常普遍尊重的话, 那就要在很大程度上归功于作为学者与导师的华罗庚五十年来对于中国数学事业的领导和贡献”. 这是国际数学界对他的赞誉, 并称他为一位奇才, 他有卓越的个人学术成就, 同时他是一位历史上罕见的发展本国数学的数学思想家、实践家.

华罗庚受到的正规教育仅到初中毕业, 之后读了一年多职业学校, 即主要靠自学成为伟大的数学家, 无疑他要比通常的数学家付出更多的辛劳. 他是我国解析数论、自守函数论、多复变函数论、典型域上的调和分析、典型群、除环论、数值分析中的数论方法及应用数学等众多领域的创始人与开拓者, 他的一些著作已经成为这些领域的经典文献.

华罗庚是一位爱国数学家. 在抗日战争刚开始时, 他即从英国回到中国, 在云南昆明执教于西南联合大学直至抗战胜利. 中华人民共和国刚成立, 他即放弃在美国伊利诺依大学的终身教授职务, 率全家于 1950 年回到祖国, 担负起领导中国数学发展的工作达三十余年, 直到去世.

华罗庚非常关心同时注意培养年轻数学家, 并能和同事们共同讨论切磋. 早在昆明时期, 受他影响而成为著名数学家的有段学复、闵嗣鹤、樊几、徐贤修等人. 1950 年以后, 受他影响与在他直接领导之下工作的人就更多了, 如冯康、越民义、万哲先、陆启铿、龚升、许孔时、王元、陈景润、丁石孙、丁夏畦、王光寅、张里千、陈希孺、吴方、魏道政、严士健、潘承洞、任建华、石钟慈、许以超、冯克勤等人. 他还为大学生写了不少教科书.

华罗庚很重视科学普及工作及数学方法在工业生产中的应用. 他是我国数学竞赛活动的创始人并为中学生写了不少课外读物. 华罗康持续了近二十年在中国各省、市、自治区普及推广工业生产中的数学方法, 给工人们讲课, 产生了很好的经济效益和社会效益. 长期跟他一起工作的有陈德泉、计雷、李之杰、徐伟宣、杨德庄等人. 他非常重视发展应用数学的探索创新工作, 与他密切合作的有王元, 受他学术思想、方法论引领与影响下工作的有曾肯成、裴定一、方开泰、杨德庄等人.

当然, 华罗庚也得到了他的前辈对他的教导与培养, 得到了他的同辈数学家对他的关心与帮助. 特别在他年轻处于最困难处境的时候, 得到了熊庆来、杨武之、叶企荪等先生对他的提拔与帮助.

我们应该全面总结华罗庚的一生, 以便后辈们能更好地以他为榜样, 将中国的数学事业搞上去, 更好地服务于祖国. 同时, 研究华罗庚无疑也是中国近代数学史研究的重要任务之一.

科学出版社邀请长期与华罗庚一起工作的几个学生负责编辑《华罗庚文集》, 无疑是一项非常有眼光的举措. 作为他的学生与晚辈, 我们都愿意积极奉献力量. 我们将首先编辑出版他的原始论文与学术专著, 将按数论、代数几何、分析、应用数学来分类编辑.

最后, 我们在此对于支持这项工作的单位与友好人士表示衷心的感谢, 特别要感谢中国科学院数学与系统科学研究院的创新资金的支持, 感谢国家出版基金的支持, 感谢中国科学院知识创新工程资助项目“中国近现代科学技术发展综合研究”(KJCX-W6) 与国家自然科学基金委“20 世纪数学史”研究项目 (2052100) 的支持. 科学出版社的编辑同志对本书的出版做了大量深入细致的工作, 在此一并感谢.

《华罗庚文集》应用数学卷编辑组

# 序

在纪念数学大师华罗庚 100 周年华诞之际, 纵观中国应用数学发展史, 从探路工作以及始于 1958 年的数学普及工作, 到创造型工作 (以闻名国际的华–王方法为代表), 到应用数学人才的培养工作, 人们清楚地看到了华罗庚和王元对中国应用数学发展的引领和推动作用, 他们是探路人和开拓者; 人们还清楚地看到他们的应用数学工作始终就密不可分. 他们是应用数学紧密合作 (团队式工作) 的典范. 因此, 出版他们的应用数学文集, 科学的、恰当的做法, 就是把华、王的工作成果放在一起出版. 这就是本书在编辑时最初确定的书名为《华罗庚、王元应用数学文集》的缘由. 但是, 王元认为他的应用数学一直都是在华老指导与影响下, 与他共同进行的, 应该属于“华罗庚应用数学体系与工作”中的一部分, 应该以华罗庚冠名出文集是最适当的, 所以本文集最后定名为《华罗庚文集: 应用数学卷》.

## 紧密合作的典范

### 1. 探路工作的密不可分

为什么要探路?

数学是什么? 应用数学是什么? 应用数学应该是什么样子? 时至今日, 国际数学界仍在争论之中. 众所周知, 尽管有争论, 但是对于纯粹数学研究而言, 国际上已有几百年来逐步形成的一种套路可循, 而应用数学则没有. 因此, 各国为发展本国的应用数学的领头人, 首先要探路 (通俗地讲, 就是应用数学搞什么? 怎么搞?).

数学研究, 领头的数学家的视点, 具有导向作用, 纯粹数学是如此 (如 Hilbert 1900 年提出的 23 个问题, 就是他汇总了前人的观点, 加他本人的感悟, 形成的一种对数学今后发展的观点, 它影响了 20 世纪的数学发展方向), 应用数学也是如此. 在探索中国应用数学之路时, 需要观点导向, 在中国应用数学起步时, 导向观点集中体现在华罗庚、王元合作的文章《有限与无穷, 离散与连续》中. 因为应用数学主要涉及数学外部的实际问题和纯粹数学与别的学科 (分支) 的交叉应用的问题, 它要构建数学模型或重构数学模型, 它还要研究对模型的数值求解的好算法. 因此,《有限与无穷, 离散与连续》的辩证统一的观点与技巧, 极为重要. 文中的观点与实例导引人们对离散性、非线性、随机性的特殊视角, 那是近年来应用数学家才充分认识到的“三性”难点, 在 20 世纪五六十年代华罗庚和王元就已点明了. 文中还强调了离散问题特殊性和离散逼近思想的重要性.

探路工作还需要寻找“问题”.

华罗庚和王元认为“问题”对纯粹数学研究和应用数学研究一样重要, 这与许多著名数学家观点一致, 比如 M. Atiyah 就强调“问题”在数学发展中起关键作用. 但是, 应用数学“问题”的寻觅与纯粹数学不同. 纯粹数学“问题”主要来自数学内部, 大量的猜想和各纯粹数学分支文献中的问题展现在读者的面前, 应用数学的问题多来自数学的外部, 寻找这样的“问题”具有相当的难度, 多数“问题”还只是自然语言的表述, 未形成数学问题; 即使与纯粹数学有关的应用数学“问题”, 没有洞察力和数学的慧敏, 也难以捕捉到. 华罗庚和王元从事应用数学工作起步时, 首先要寻觅应用数学“问题”. 他们毕竟首先是纯粹数学家, 在中国对应用数学需求先天不足的那个时代, 他们只好先从书本上、文献上找“问题”. 他们成功了, 他们找到了“问题”. 他们在寻找“问题”的工作上, 就密不可分, 那是一个个过程, 是一个个有趣的故事. 这里叙述两个故事. 华罗庚想到采矿与水利等方面可能有数学问题, 于是他让王元到北京各有关部门去了解情况. 王元在北京矿业学院的教师那里借来了几本“矿体几何学”的书. 华罗庚从他的朋友陆漱芬那里学到了地理学家计算坡地表面积的方法. 这样结合起来, 他们共同找到了应用数学的第一个“问题”, 研究完成了第一篇应用数学论文. 第二个故事是“华–王方法”的“问题”, 关于数论在多重积分近似计算中的应用问题. 苏联是世界数学强国, 1957 年苏联科学院的工作总结中提到了两项重要数学成果, 两项之一为数论在多重积分近似计算中的应用. 有一篇俄文的文章中讲到积分近似计算中的蒙特卡罗方法, 讲到其中所需的随机数服从一致分布等. 王元拿了文章去找华罗庚谈. 那天华罗庚很累, 不想看. 王元说:“就看这一行, 行不行?”华罗庚看后很兴奋地说:“蒙特卡罗方法实质上就是数论中的一致分布论的应用, 这就好像隔着一层纸, 戳穿了就那么一点点东西.”此时他们俩心有灵犀一点通, 已经共同洞察到这个问题更深层的数学现象. 他们立即捕捉住他们共同寻觅的“问题”. 这是他们共同寻觅的第二个“问题”.

**2. 创造型攻关研究的紧密合作.**

纯粹数学研究成果主要靠纯粹数学家个体完成的. 比如, 华罗庚、王元、陈景润在数论领域的辉煌成就, 都是靠他们个人的聪慧才智登上了别人达不到的高峰. 当然也有特例, 比如, “对于整整一代人来说, 哈代、利特伍德的合作主宰了英国的纯粹数学, 也在很大程度上主宰了世界的解析数论. 但是至今没有人知道他们是如何合作的”. 应用数学攻关研究要靠团队的合作力量, 华罗庚、王元从事应用数学研究一开始，就给后来人树立了榜样. 他们对共同寻找到的应用数学问题进行攻关研究时, 紧密合作. 对于第一个“问题”, 华罗庚首先证明了一个漂亮的不等式

$$V \leqslant B \leqslant S$$

($V$、$B$、$S$ 分别表示地理学家伏尔柯夫方法、矿业学家包曼方法求坡面面积的极限结果, $S$ 是坡面的真面积).

王元用华罗庚的方法对"矿体几何学"书上介绍的估计矿床体积的方法进行理论探讨, 得到了一系列好结果. 他们一起只用一点微积分, 只用了 12 页, 就把 300 多页的"矿体几何学"上的计算方法全写出来了. "关于在等高线图上计算矿藏储量与坡地面积的问题", 就是他们共同攻关的第一个"问题"的成果.

对于第二个"问题", 华罗庚、王元密切合作进行攻关研究就更精彩了. 就在 1958 年, 华罗庚与王元在苏联的《科学通报》上见到数论方法用于高维数值积分的第一篇理论文章, 他们有特殊的敏感性和洞察力, 看到了更深层的问题, 他们立即开展深入研究. 他们从 2 维入手, 华罗庚猜出用斐波那契数列来构造 2 维一致分布点列的方法可以得到最佳求积公式, 王元只用两页纸就证明了一个重要的公式, 虽然同时代其他数学家也得到同样的结果, 但所用的方法要间接而麻烦得多.

接着, 他们要转向高维空间, 华罗庚凭他特有的数学直觉, 从斐波那契数列中相邻两数之比是黄金数 $\frac{\sqrt{5}-1}{2}$ 的渐近分数出发, 又提出理想的思路. 但高维问题逻辑推理遇到了困难. 他们仍然紧密合作进行攻关研究, 有半年多时间, 王元一清早就去华罗庚家, 在他家一起进早餐, 饭后就演算, 但仍无进展, 攻关陷入困境. 后来新的转机出现了, 是历史的巧合, 也是播种与收获. 他们的攻关研究在中国第一台电子管电子计算机上算出结果. 华罗庚正是中国电子计算机研制的倡导者和组织者, 由于他的工作, 中国才有了第一台电子管计算机. 攻关成功还在于王元的另一个创造性思维, 他借助文献的启发, 放弃了逻辑推导来证明定理的手段, 改用计算机模拟的手段, 即根据华罗庚关于分圆域的想法, 编了一个计算程序, 先在台式计算器上算出点列, 然后请计算所的人在电子计算机上算出这个点列对应的求积公式的最大误差, 最终在电子计算机上算出了结果. 这是华罗庚、王元的密切合作的成果, 它是一个"构造性的方法". 计算量是 $\log n$ 数量级, 而一般的计算量是 $n^2$ 数量级. 他们把结果发表在《中国科学》的《研究简报》栏上. "文革"爆发后, 他们的工作中断了. 至 1972 年, 华罗庚参加了由廖承志率领的一个代表团访问了日本. 日本数学家告诉他, "华–王方法"很成功, 并在有关文献中看到了首次以"华–王方法"来命名他们的成果. "文革"结束前夕, 他们将研究结果写成几篇论文发表在《中国科学》上, 并着手撰写专著《数论在近似分析中的应用》, 全面总结了这一领域的成就, 该书于 1978 年由科学出版社出版. 1981 年, 施普林格出版社与科学出版社联合出版了这本书的英文版.

我们将两位数学家华罗庚、王元关于"华–王方法"陆续发表的论文, 以及施普林格出版社与科学出版社联合出版英文版《数论在近似分析中的应用》一并编辑出版. 做出范式以便后来者评读.

用计算机模拟代替逻辑推导来证明定理的手段, 不但产生了"华–王方法", 而

且在中国可能是最早成功的范例. 计算机模拟技术, 后来发展为计算机仿真技术, 其核心是数学技术, 俄罗斯数学家 A. A. Samarskii 称其为“数值实验”方法, 并称其为一种新的科学方法.

“华–王方法”的产生, 还给人们一个启示：灵活性的重要性, 假若华罗庚、王元坚持用逻辑推导来证明定理的手段不放弃, 可能要很长时间才出成果, 而且还失去了创造性的构造性方法的产生. 灵活地转换手法和途径, 是应用数学家最重要的素质之一.

华罗庚与王元的密切合作, 完全不同于哈代与利特伍德的合作. 他们的数学思想、方法技巧的结合, 我们可以从王元的相关回忆文章中得到更深刻的感悟.

这项工作能够顺利完成是华罗庚与王元密切合作的结果, 王元非常感谢华罗庚对他的指导, 并屡次提出宝贵的原始数学思想; 华罗庚也对王元提出来要研究这一课题而感到满意, 他在一张字条上谦虚地写道“被王元拉上一条路”, 又写道“我对蒙特卡罗方法的一知半解, 就是在年轻人帮助之下学来的”, 真是“多年师生成兄弟, 共同学习共钻研”.

“一些对华罗庚了解不深的人往往以为他的最大优点是逻辑推导与计算能力强, 其实他最强的数学才能恰好是他的数学直觉. 华罗庚的另一个特点是先从一个具体而简单的特例着手研究的单刀直入式的研究方式.”

## 两个层面的工作, 两种不同的成果

华罗庚、王元在中国发展应用数学, 一直是在两个层面上开展工作：一个层面是普及型, 另一个层面是创造型. 普及型工作分成两类：一类是面向中学生的, 另一类是面向大众的. 面向大众的普及型工作和创造型工作都始于 1958 年. 那一年, 全国首次推广运筹学中的线性规划方法 —— 中国独特的“图上作业法”, 曾形成群众运动, 并在山东济南召开过现场会; 那一年, 在应用数学人才培养方面, 在新创建的中国科学技术大学, 设立“应用数学与计算技术系”(这是在高校首次开办应用数学系), 华罗庚、王元从基础课教学开始, 在中国科大培养应用数学人才; 还在那一年, 华罗庚、王元探索创造型层面的研究也开始了. 就在那一年, 他们找到了后来世称“华–王方法”的“问题”, 从找到“问题”并立即投入研究, 持续到 1978 年科学出版社出版专著, 再到 1981 年施普林格和科学出版社联合出版专著的英文版, 一系列高水平成果遍布在这 20 多年期间 (尽管“文革”中中断了几年); 另一项创造型的研究 —— 数学在国民经济中的应用, 同一时间也开始了, 有关思想和成果在“有限与无穷, 离散与连续”中已有展示, 这项工作被盗毁于“文革”期间, 华罗庚在 1981 年开始又重新回忆并加上新的创造形成新的成果, 王元又对这个成果进行整理改写, 到此为止, 也时续了 20 多年.

华罗庚的前期数学普及工作, 从 1958 年开始到 1960 年, 除了普及“线性规划”外, 还有农村“麦场设置”, 以他名义曾在《光明日报》和《数学学报》上发表过有关文章, 王元与万哲先执笔写了《物资调运工作中的数学方法》一书, 王元与朱永津等又执笔撰写了一本教科书《线性规划的理论及其应用》. 这些都未收录到本文集中. 华罗庚的后期数学普及工作始于 1965 年, 以普及推广统筹法、优选法为主要内容. 统筹法、优选法也是从书本、文献中选出来的, 但与创造型不同, 创造型挑选的是“问题”, 普及型挑选的是“方法”, 是“技术”, 是可以加工成通俗易懂的方法或技术. 华罗庚在“学术上洞察之深、选材之妙、加工之巧、表达之深入浅出”, 他写的普及读本, 真正让大众看得懂、学得会、用得上、见成效. 在编辑这本文集时, 普及型的文集首选的是两个评话, 它是普及著作的精品, 是范本.

华罗庚面向中学生的教学普及著作精品已有专辑出版. 一般说来, 它不属于应用数学文集范围, 本文集不再重复刊登. 但有一文例外, “谈谈与蜂房结构有关的数学问题”, 它既是给中学生讲的数学普及读物, 又是从刊物上找到的应用数学“问题”的研究成果. 蜂房的优化结构、生物现象、自然界奇迹, 其中最迷人的是数学现象. 小小的蜜蜂怎么解决蜂房优化结构的数学问题. 华罗庚的“始之以有趣”、“好奇”, 使他放不下在通俗读物上看到的奇观, 他抓住了这个“问题”, 展开研究. 怎么展开, 怎么深入, 怎么……, 只要看他写的十六个小标题, 任何人都会被吸引住 (0 楔子、一有趣、二困惑、三访实、四解题、五浅化、六慎微、七切方、八疑古、九正题、十设问、十一代数、十二几何、十三推广、十四极限、十五抽象). 这又是一篇创造型的范文.

应用统计中的数论方法, “问题”来自实际 —— 导弹设计中提出的试验设计问题. 王元、方开泰提出的“均匀设计”是具有独创性的成果. “追本溯源, 若无华罗庚对近似分析中数论方法的倡导与工作, 很难设想这项工作能在中国这样快地发展起来, 所以也应该部分地归功于华罗庚”. 本文集选用了这领域的几篇最重要的, 与数学关系较多原创性文章.

华罗庚曾写过《数学方法与国民经济》一书的征求意见初稿, 该书分三部分, 用“前言”、“中论”、“后语”分开. 该书初稿中提到“在本世纪 (指 20 世纪) 中叶……想把国民经济提上去的愿望, 明知学识和经验不足, 宁可放着驾轻就熟的理论专长于第二位, 硬着头皮进行尝试, 初步归纳出 12 个字: 大统筹, 广优选, 联运输, 策发展. 再经过发展, 又提出 36 字: 大统筹, 广优选, 联运输, 精统计, 抓质量, 理数据, 建系统, 策发展, 利工具, 巧计算, 重实践, 明真理.

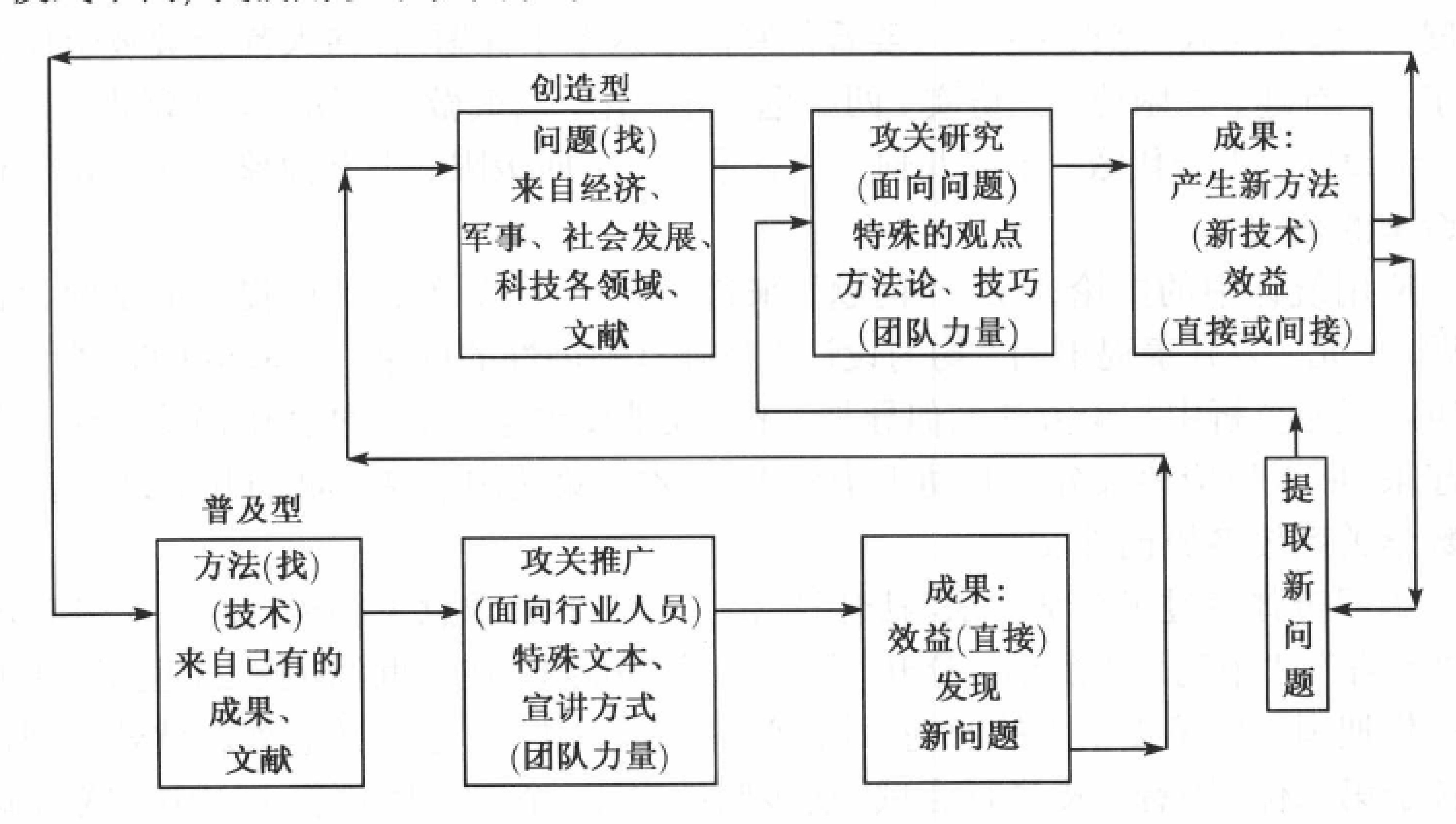
从 12 字发展到 36 字. 以建系统, 策发展的“正特征矢量法”的优化经济数学理论与技术为终结. 该书未经作者审定出版, 本文集也未把这部分内容收集在内.

华罗庚在评论一个人的工作时, 曾说过“一个人的工作有几项, 比如讲有两三项, 在历史上留下来就很了不起. 一个人一辈子发表了几百篇论文, 许多著作, 真正

能在历史上记一笔的就那么几项, 其他就随风飘了". 他在谈自己工作时, 在应用数学方面, 就提到了两项: 其一就是数论在近似分析中的应用 (华–王方法); 其二是经济系统优化方法, 并且他把它和普及统筹法、优选法放在一起. 他说从 60 年代初开始, 为中国经济建设服务的应用数学工作, 主要是建了一个"门", "门"字的两竖杠是两根柱子: 一边是"统筹法", 另一边是"优选法"; "门"的横梁是"正特征矢量法". 他明确指出, 统筹法和优选法可以作为他经济系统优化理论与方法 (技术) 的基础性方法 (技术).

本文集仅收集华罗庚、王元的应用数学方面已发表的论文, 不包含他们纯粹数学方面的论文. 即使应用数学方面, 也不是他们的全部工作的反映, 这与纯粹数学不同, 因为应用数学的许多工作是不能以论文形式发表的. 再者, 我们也不做他们的成果评价, 只不过应用数学特殊性 (人们对它的认识等方面), 我们不得不多花了一些笔墨去描述, 比如探路工作、工作特点、普及型、创造型等等; 但是, 从文集的文章还不能全面反映他们应用数学观点和方法论特色, 比如数学现象、数学技术、数学工程、模型论、算法论、团队论、交叉综合论等观点. 请见附录 2、3.

两个层面的工作和两种不同的成果, 还要补充几句话. 首先两者工作过程和工作模式不同, 我们用以下框图表示:




框图表示意在给人以明快、直观的逻辑思维模式和整体的工作流程. 从框图可见:

创造型研究, 始于"问题", 经过奋力攻关, 形成的成果是产生新的数学技术 (新的数学方法), 同时产生直接或间接的效益.

例如,

① 始于"数论在近似分析中的应用"的问题, 奋力攻关后, 产生了"华–王方

法”(新的数学技术). 应用于各个领域, 产生了效益, 也具有很高的学术水平.

② 始于中国经济背景的“经济优化发展”问题, 奋力攻关后, 产生了“正特征矢量法”(新的数学技术). 它与列昂铁夫的“投入产生法”不同:“投入产生法”意在经济系统的平衡; “正特征矢量法”旨在经济系统的优化发展, 给出优化发展的策略. “正特征矢量法”, 虽未有实践, 那是因为各种条件限制以及认知不足, 终究可能会被采用的, 华罗庚的学生们一直在为此而努力. 即使在西方经济, 社会发展条件下, 列昂铁夫的“投入产出法”(1936 年提出) 也经过了 11 年 (1947 年) 才在实际中列出第一张投入产出表.

普及推广型, 始于数学技术 (数学方法), 经过成功的加工运作, 努力普及推广, 产生直接 (或间接) 的经济社会效益.

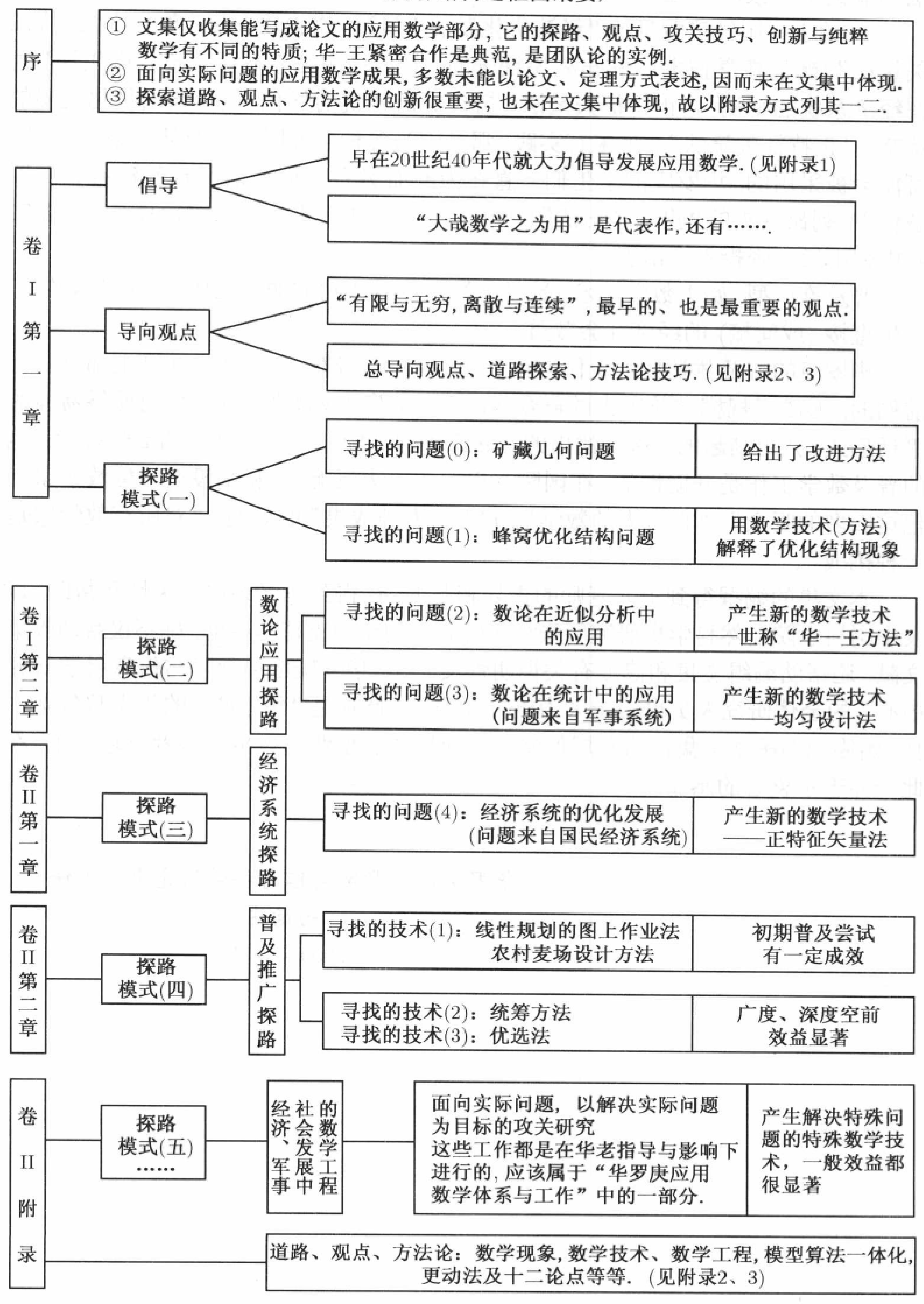
华罗庚的普及数学技术工作与众不同, 极具创造性. 在广度、深度上都是史无前例的, 形成了规模空前宏大的群众运动. 他在普及文本加工时, 采用通俗易懂的平话形式, 并用高超的、深入浅出的、形象化的讲授方式向大众介绍数学技术. 他的普及数学工作是开创性的, 在国际上引起了极大反响, “从来没有一位数学家有他这么多的听众”、“百万人的数学”, 产生“万项成果”的效益, “对所有数学家是一种挑战”.

本文集的编撰得到王元老师的指导和帮助, 还得到华老亲属的关怀和帮助, 以及中国科学院自然科学史研究所张利华研究员的全力支持和帮助, 她不仅帮助收集文献, 还帮助编辑文集和修改有关说明性文章, “华罗庚应用数学与信息科学研究中心”的全体研究人员也给予大力支持和帮助, 特别是“华中心”的华光常务副主任, 始终把编撰本文集作为发展他父亲的应用数学事业的大事与编者一起工作, 在此一并表示衷心的感谢.

华罗庚应用数学与信息科学研究中心主任
杨德庄
2009 年 8 月

# 华罗庚文集应用数学卷

（内容结构之框图纲要）



# 目　　录

**第一章　倡导、导向观点、探路 (I) 之代表作** …… 1

- 大哉数学之为用 …… (华罗庚) 3
- 有限与无穷, 离散与连续 …… (华罗庚、王元) 13
- 关于在等高线图上计算矿藏储量与坡地面积的问题 …… (华罗庚、王元) 41
- 谈谈与蜂房结构有关的数学问题 …… (华罗庚) 56

**第二章　创造型工作 (I)、探路 (II) 之代表作** …… 87

甲. 近似分析中的数论方法 …… (华罗庚、王元) 89

- 关于多重积分的近似计算的若干注记 …… (华罗庚、王元) 91
- 某类函数插值公式的一个注记 …… (王元) 95
- 丢番图逼近与数值积分 (I) …… (华罗庚、王元) 100
- 丢番图逼近与数值积分 (II) …… (华罗庚、王元) 104
- 多维周期函数的数值积分 …… (华罗庚、王元) 108
- 关于一类函数的插入公式 …… (王元) 123
- 论一致分布与近似分析 —— 数论方法 (I) …… (华罗庚、王元) 127
- 论一致分布与近似分析 —— 数论方法 (II) …… (华罗庚、王元) 153
- 论一致分布与近似分析 —— 数论方法 (III) …… (华罗庚、王元) 174
- Applications of Number Theory to Numerical Analysis …… (华罗庚、王元) 190

乙. 应用统计中的数论方法 …… (王元、方开泰) 421

- 关于均匀分布与试验设计 (数论方法) …… (王元、方开泰) 423
- 应用统计中的数论方法(I) …… (王元、方开泰) 430
- 应用统计中的数论方法 (II) …… (王元、方开泰) 448
- 混料均匀设计 …… (王元、方开泰) 460
- 统计模拟中的数论方法 …… (王元、方开泰) 472

# 第一章

# 倡导、导向观点、探路(Ⅰ)之代表作


- 大哉数学之为用……………………………………………(华罗庚) 3
- 有限与无穷，离散与连续…………………(华罗庚、王元) 13
- 关于在等高线图上计算矿藏储量与坡地面积的问题…………………………(华罗庚、王元) 41
- 谈谈与蜂房结构有关的数学问题…………………(华罗庚) 56

# 大哉数学之为用[①]

## 数 与 量

数 (读作 shù) 起源于数 (读作 shǔ), 如一、二、三、四、五 ⋯⋯, 一个、两个、三个 ⋯⋯. 量 (读作 liàng) 起源于量 (读作 liáng). 先取一个单位作标准, 然后一个单位一个单位地量. 天下虽有各种不同的量 (各种不同的量的单位如尺、斤、斗、秒、伏特、欧姆和卡路里等等), 但都必须通过数才能确切地把实际的情况表达出来. 所以"数"是各种各样不同量的共性, 必须通过它才能比较量的多寡, 才能说明量的变化.

"量"是贯穿到一切科学领域之内的, 因此数学的用处也就渗透到一切科学领域之中. 凡是要研究量、量的关系、量的变化、量的关系的变化、量的变化的关系的时候, 就少不了数学. 不仅如此, 量的变化还有变化, 而这种变化一般也是用量来刻画的. 例如, 速度是用来描写物体的变化的动态的, 而加速度则是用来刻画速度的变化. 量与量之间有各种各样的关系, 各种各样不同的关系之间还可能有关系. 为数众多的关系还有主从之分 —— 也就是说, 可以从一些关系推导出另一些关系来. 所以数学还研究变化的变化, 关系的关系, 共性的共性, 循环往复, 逐步提高, 以至无穷.

数学是一切科学得力的助手和工具. 它有时由于其他科学的促进而发展, 有时也先进一步, 领先发展, 然后再获得应用. 任何一门科学缺少了数学这一项工具便不能确切地刻画出客观事物变化的状态, 更不能从已知数据推出未知的数据来, 因而就减少了科学预见的可能性, 或者减弱了科学预见的精确度.

恩格斯说: "纯数学的对象是现实世界的空间形式和数量关系". 数学是从物理模型抽象出来的, 它包括数与形两方面的内容. 以上只提要地讲了数量关系, 现在我们结合宇宙之大来说明空间形式.

## 宇 宙 之 大

宇宙之大, 宇宙的形态, 也只有通过数学才能说得明白. 天圆地方之说, 就是古代人民用几何形态来描绘客观宇宙的尝试. 这种"苍天如圆盖, 陆地如棋局"的宇

① 本文曾于 1959 年 5 月 28 日发表在"人民日报"上. 后曾以"数学的用场与发展"为题转载在《现代科学技术简介》(科学出版社,1978 年) 上. 转载时, 作者认为时代已有很大发展, 内容要重新修改补充. 由于时间仓促, 只能根据他的口述笔录对原稿加以整理发表. 他再三提出, 希望听取各方面的宝贵意见, 以便在适当时候对这篇文章加以补充修改.

宙形态的模型, 后来被航海家用事实给以否定了. 但是, 我国从理论上对这一模型提出的怀疑要早得多, 并且也同样地有力. 论点是:“混沌初开, 乾坤始奠, 气之轻清上浮者为天, 气之重浊者下凝者为地. ”但不知轻清之外, 又有何物? 也就是圆盖之外, 又有何物? 三十三天之上又是何处? 要想解决这样的问题, 就必须借助于数学的空间形式的研究.

四维空间听来好像有些神秘, 其实早已有之, 即以“宇宙”二字来说, “往古来今谓之宙, 四方上下谓之宇”(《淮南子 · 齐俗训》) 就是宇是东西、南北、上下三维扩展的空间, 而宙是一维的时间. 牛顿时代对宇宙的认识也就是如此. 宇宙是一个无边无际的三维空间, 而一切的日月星辰都安排在这框架中运动. 找出这些星体的运动规律是牛顿的一大发明, 也是物理模型促进数学方法, 而数学方法则是用来说明物理现象的一个好典范. 由于物体的运动不是等加速度, 要描绘不是等加速度, 就不得不考虑速度时时在变化的情况, 于是乎微商出现了. 这是刻画加速度的好工具. 由牛顿当年一身而二任焉, 既创造了新工具 —— 微积分, 又发现了万有引力定律. 有了这些, 宇宙间一切星辰的运动初步统一地被解释了. 行星凭什么以椭圆轨道绕日而行的, 何时以怎样的速度达到何处等, 都可以算出来了.

有人说西方文明之飞速发展是由于欧几里得几何的推理方法和进行系统实验的方法. 牛顿的工作也是逻辑推理的一个典型. 他用简单的几条定律推出整个的力学系统, 大至解释天体的运行, 小到造房、修桥、杠杆、称物都行. 但是人们在认识自然界时建立的理论总是不会一劳永逸完美无缺的, 牛顿力学不能解释的问题还是有的. 用它解释了行星绕日公转, 但行星自转又如何解释呢? 地球自转一天 24 小时有昼有夜, 水星自转周期和公转一样, 半面永远白天, 半面永远黑夜. 一个有名的问题: 水星进动每百年 42″, 是牛顿力学无法解释的.

爱因斯坦不再把“宇”、“宙”分开来看, 也就是时间也在进行着. 每一瞬间三维空间中的物质在占有它一定的位置. 他根据麦克斯韦 - 洛伦兹的光速不变假定, 并继承了牛顿的相对性原理而提出了狭义相对论. 狭义相对论中的洛伦兹变换把时空联系在一起, 当然并不是消灭了时空特点. 如向东走三里, 再向西走三里, 就回到原处, 但时间则不然, 共用了走六里的时间. 时间是一去不复返地流逝着. 值得指出的是有人推算出狭义相对论不但不能解释水星进动问题, 而且算出的结果是“退动”. 这是误解. 我们能算出进动 28″, 即客观数的三分之二. 另外, 有了深刻的分析, 反而能够浅出, 连微积分都不要用, 并且在较少的假定下, 就可以推出爱因斯坦狭义相对论的全部结果.

爱因斯坦进一步把时、空、物质联系在一起, 提出了广义相对论, 用它可以算出水星进动是 43″, 这是支持广义相对论的一个有力证据, 由于证据还不多, 因此对广义相对论还有不少看法, 但它的建立有赖于数学上的先行一步. 如先有了黎曼几何. 另一方面它也给数学提出了好些到现在还没有解决的问题. 对宇宙的认识还将

有多么大的进展，我不知道，但可以说，每一步都是离不开数学这个工具的.

## 粒子之微

佛经上有所谓"金粟世界"，也就是一粒粟米也可以看作一个世界. 这当然是佛家的幻想. 但是我们今天所研究的原子却远远地小于一粒粟米，而其中的复杂性却不亚于一个太阳系.

即使研究这样小的原子核的结构也还是少不了数学. 描述原子核内各种基本粒子的运动更是少不了数学. 能不能用处理普遍世界的方法来处理核子内部的问题呢？情况不同了！在这里，牛顿的力学，爱因斯坦的相对论都遇到了困难. 在目前人们应用了另一套数学工具. 如算子论，群表示论，广义函数论等. 这些工具都是近代的产物. 即使如此，也还是不能完整地说明它.

在物质结构上不管分子论、原子论也好，或近代的核子结构、基本粒子的互变也好，物理科学上虽然经过了多次的概念革新，但自始至终都和数学分不开. 不但今天，就是将来，也有一点是可以肯定的，就是一定还要用数学.

是否有一个统一的处理方法，把宏观世界和微观世界统一在一个理论之中，把四种作用力统一在一个理论之中，这是物理学家当前的重大问题之一. 不管将来他们怎样解决这个问题，但是在处理这些问题的数学方法必须统一. 必须有一套既可以解释宏观世界又可以解释微观世界的数学工具. 数学一定和物理学刚开始的时候一样，是物理科学的助手和工具. 在这样的大问题的解决过程中，也可能如牛顿同时发展天体力学和发明微积分那样，促进数学的新分支的创造和形成.

## 火箭之速

在今天用"一日千里"来形容慢则可，用来形容快则不可了！人类可创造的物体的速度远远地超过了"一日千里". 飞机虽快到日行万里不夜，但和宇宙速度比较，也显得缓慢得很. 古代所幻想的朝昆仑而暮苍梧，在今天已不足为奇.

不妨回忆一下，在星际航行的开端 —— 由诗一般的幻想进入科学现实的第一步，就是和数学分不开的. 早在牛顿时代就算出了每秒钟近八公里的第一宇宙速度，这给科学技术工作者指出了奋斗目标. 如果能够达到这一速度，就可以发射地球卫星. 1970 年我国发射了第一颗人造卫星. 数学工作者自始至终都参与这一工作 (当然，其中不少工作者不是以数学工作者见称，而是运用数学工具者). 作为人造行星环绕太阳运行所必须具有的速度是 11.2 公里/秒，称为第二宇宙速度；脱离太阳系飞向恒星际空间所必须具有的速度是 16.7 公里/秒，称为第三宇宙速度. 这样的目标，也将会逐步去实现.

顺便提一下, 如果我们宇宙航船到了一个星球上, 那儿也有如我们人类一样高级的生物存在. 我们用什么东西作为我们之间的媒介. 带幅画去吧, 那边风景殊, 不了解, 带一段录音去吧, 也不能沟通. 我看最好带两个图形去. 一个“数”, 一个“数形关系”(勾股定理)(图 1 和图 2)

为了使那里较高级的生物知道我们会几何证明, 还可送去下面的图形, 即“青出朱入图”(图 3). 这些都是我国古代数学史上的成就.

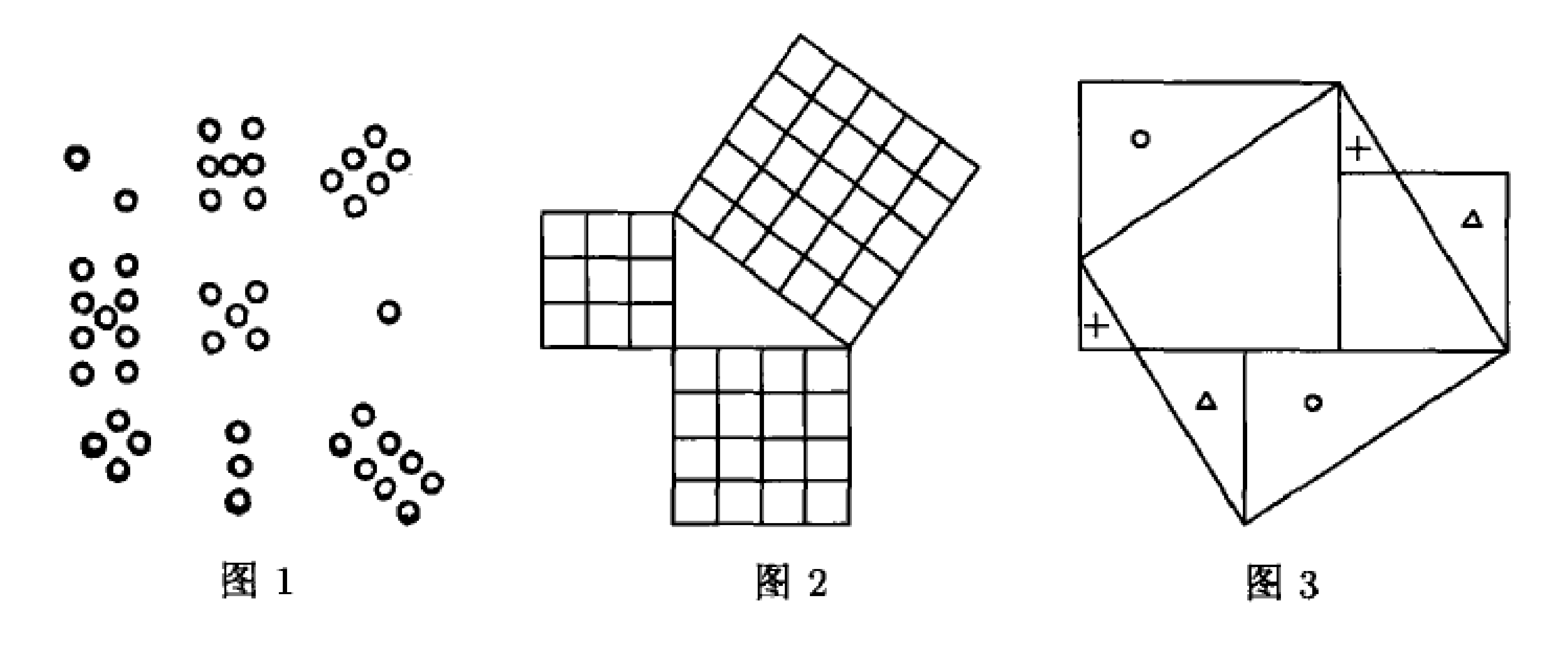


图 1　　图 2　　图 3

# 化工之巧

化学工业制造出的千千万万种新产品, 使人类的物质生活更加丰富多彩, 真是“巧夺天工”, “巧夺造化之工”. 在制造过程中, 它的化合与分解方式是用化学方程来描述的, 但它是在变化的, 因此, 伟大革命导师恩格斯明确指出:“表示物体的分子组合的一切化学方程式, 就形式来说是微分方程式. 但是这些方程式实际上已经由于其中所表示的原子量而积分起来了. 化学所计算的正是量的相互关系为已知的微分.”

为了形象化地说明, 例如, 某种物质中含有硫, 用苯提取硫. 苯吸取硫有一定的饱含量, 在这个过程中, 苯含硫越多越难再吸取硫, 剩下的硫越少越难被苯吸取. 这个过程时刻都在变化, 吸收过程速度在不断减慢着. 实验本身便是这个过程的积分过程, 它的数学表达形式就是微分方程式及其求解. 简单易作的过程我们可以用实验去解决, 但对于复杂、难作的过程, 则常常需要用数学手段来加以解决. 特别是选取最优过程的工艺, 数学手段更成为必不可少的手段. 特别是量子化学的发展, 使得化学研究提高到量子力学的阶段, 数学手段 —— 微分方程及矩阵、图论更是必需的数学工具.

应用了数学方法还可使化学理论问题得到极大的简化. 例如, 对于共轭分子的能级计算, 在共轭分子增大时十分困难. 应用了分子轨道的图形理论, 由图形来简化计算, 取得了十分直观和易行的效果, 便是一例, 其主要根据是如果一个行列式

中的元素为 0 的多, 那就可以用图论来简化计算.

## 地 球 之 变

我们所生活的地球处于多变的状态之中, 从高层的大气, 到中层的海洋, 下到地壳, 深入地心都在剧烈地运动着, 而这些运动规律的研究也都用到数学.

大气环流, 风云雨雪, 天天需要研究和预报, 使得农民可以安排田间农活, 空中交通运输可以安排航程. 飓风等灾害性天气的预报, 使得海军、渔民和沿海地区能够及早预防, 减少损害. 而所有这些预报都离不了数学.

"风乍起, 吹皱一池春水." 风和水的关系自古便有记述, "无风不起浪". 但是风和浪的具体关系的研究, 则是近代才逐步弄清的, 而在风与浪的关系中用到了数学的工具, 例如偏微分方程的间断解的问题.

大地每年有上百万次的地震, 小的人感觉不到, 大的如果发生在人烟稀少的地区, 也不成大灾. 但是每年也有几次在人口众多的地区的大震, 形成大灾. 对地壳运动的研究, 对地震的预报, 以及将来进一步对地震的控制都离不开数学工具.

## 生 物 之 谜

生物学中有许许多多的数学问题. 蜜蜂的蜂房为什么要像如下的形式 (图 4), 一面看是正六角形, 另一面也是如此. 但蜂房并不是六棱柱, 而它的底部是由三个菱形所拼成的.

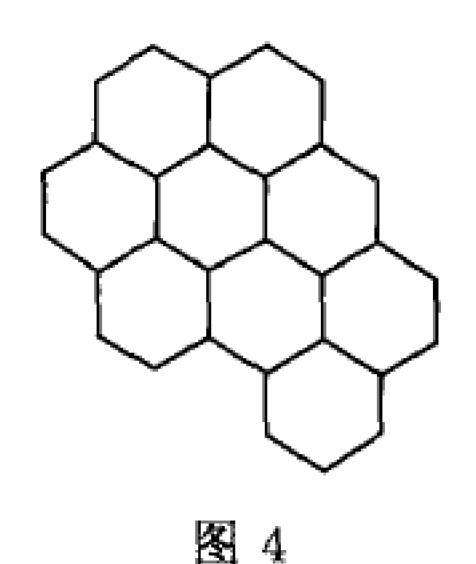


图 4

图 5 是蜂房的立体图. 这个图比较清楚, 更具体些, 拿一支六棱柱的铅笔未削之前, 铅笔一端形状是 $ABCDEF$ 正六角形 (图 6). 通过 $AC$, 一刀切下一角, 把三角形 $ABC$ 搬置 $AOC$ 处. 过 $AE, CE$ 也如此同样切三刀, 所堆成的形状就是图 7, 而蜂巢就是两排这样的蜂房底部和底部相接而成.

关于这个问题有一段趣史: 巴黎科学院院士数学家克尼格, 从理论上计算, 为使消耗材料最少, 菱形的两个角度应该是 $109^\circ 26'$ 和 $70^\circ 34'$. 与实际蜜蜂所做出的仅相差 2 分. 后来苏格兰数学家马克劳林重新计算, 发现错了的不是小小的蜜蜂,

而是巴黎科学院的院士, 因克尼格用的对数表上刚好错了一个字. 这十八世纪的难题, 1964 年我用它来考过高中生, 不少高中生提出了各种各样的证明.

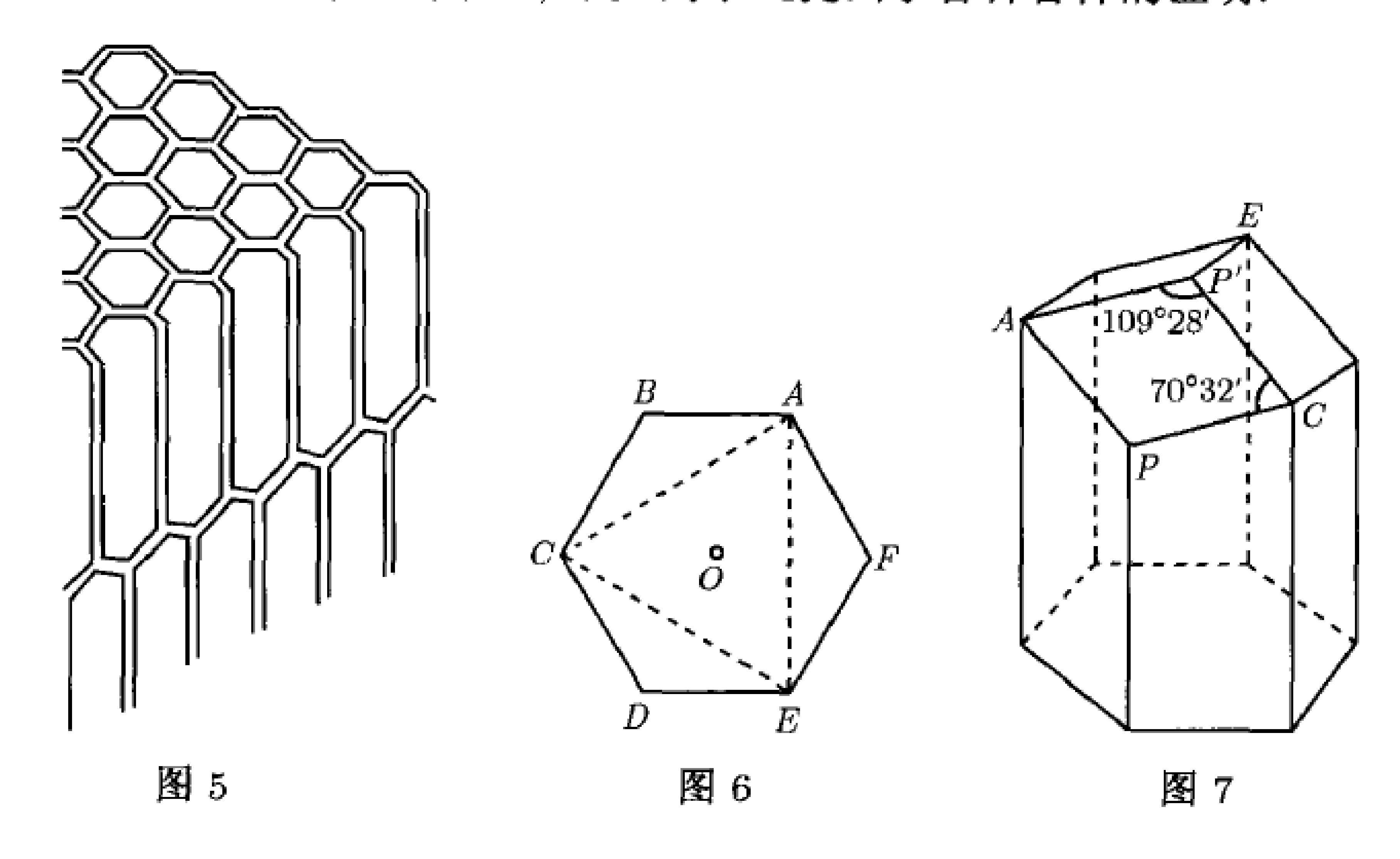


图 5　　图 6　　图 7

这一问题, 我写得篇幅略长些, 目的在于引出生物之谜中的数学, 另一方面也希望生物学家给我们多提些形态的问题, 蜂房与结晶学联系起来, 这是“透视石”的晶体.

再回到化工之巧, 有多少种晶体可以无穷无尽、无空无隙地填满空间, 这又要用到数学. 数学上已证明, 只有 230 种.

还有如胰岛素的研究中, 由于复杂的立体模型也用了复杂的数学计算. 生物遗传学中的密码问题是研究遗传与变异这一根本问题的, 它的最终解决必然要考虑到数学问题. 生物的反应用数学加以描述成为工程控制论中“反馈”的泉源. 神经作用的数学研究为控制论和信息论提供了现实的原型.

## 日用之繁

日用之繁, 的确繁, 从何谈起真为难! 但也有容易处. 日用之繁与亿万人民都有关, 只要到群众中去, 急工农之所急, 急生产和国防之所急, 不但可以知道哪些该搞, 而且知道轻重缓急. 群众是真正的英雄, 遇事和群众商量, 不但政治上有提高, 业务上也可以学到书本上所读不到的东西. 像我这样自学专攻数学的, 也在各行各业师傅的教育下, 学到了不少学科的知识, 这是一个大学一个专业中所学不到的.

我在日用之繁中搞些工作始于 1958 年, 但真正开始是 1964 年接受毛主席的亲笔指示后. 并且使我永远不会忘记的是在我刚迈出一步写了《统筹方法平话》下到基层试点时, 毛主席又为我指出:“不为个人, 而为人民服务, 十分欢迎”的奋斗目标. 后来在周总理关怀下又搞了《优选法》. 由于各省、市、自治区的领导的关怀,

我曾有机会到过二十个省市、下过数以千计的工矿农村, 拜得百万工农老师, 形成了有工人、技术人员和数学工作者参加的普及、推广数学方法的小分队. 通过群众性的科学实验活动证明, 数学确实大有用场, 数学方法用于革新挖潜, 能为国家创造巨大的财富. 回顾已往, 真有“抱着金饭碗讨饭吃”之感.

由于我们社会主义制度的优越性, 在这一方面可能有我们自己的特点, 不妨结合我下去后的体会多谈一些.

统筹方法不仅可用于一台机床的维修、一所房屋的修建、一组设备的安装、一项水利工程的施工, 更可用于整个企业管理和大型重点工程的施工会战. 大庆新油田开发, 万人千台机的统筹, 黑龙江省林业战线采、运、用、育的统筹, 山西省大同市口泉车站运煤统筹, 太原铁路局太钢和几个工矿的联合统筹, 还有一些省市公社和大队的农业生产统筹等等, 都取得了良好效果. 看来统筹方法宜小更宜大. 大范围的过细统筹效果更好, 油水更大. 特别是把方法交给广大群众, 结合具体实际, 大家动手搞起来, 由小到大、由简到繁, 在普及的基础上进一步提高, 收效甚大. 初步设想可以概括成十二个字: 大统筹, 理数据, 建系统, 策发展, 使之发展成一门学科——统筹学, 以适应我国具体情况, 体现我们社会主义社会特点. 统筹的范围越大, 得到和用到的数据也越多. 我们不仅仅是消极地统计这些数据, 而且还要从这些数据中取出尽可能多的信息来作为指导. 因此数据处理提到了日程上来. 数据纷繁就要依靠电子计算机. 新系统的建立和旧系统的改建和扩充, 都必须在最优状态下运行. 更进一步就是策发展, 根据今年的情况明年如何发展才更积极又可靠, 使国民经济的发展达到最大可能的高速度.

优选法是采用尽可能少的试验次数, 找到最好方案的方法. 优选学作为这类方法的数学理论基础, 已有初步的系统研究. 实践中, 优选法的基本方法, 已在大范围内得到推广. 目前, 我国化工、电子、冶金、机械、轻工、纺织、交通、建材等等方面都有较广泛的应用. 在各级党委的领导下, 大搞推广应用优选法的群众活动, 各行各业搞, 道道工序搞, 短期内就可以应用优选法开展数以万计项目的试验. 使原有的工艺水平普遍提高一步. 在不添人、不增设备、不加或少加投资的情况下, 就可收到优质、高产、低耗的效果. 例如, 小型化铁炉, 优选炉形尺寸和操作条件, 可使焦铁比一般达 1:18. 机械加工优选刀具的几何参数和切削用量, 工效可成倍提高. 烧油锅炉, 优选喷枪参数, 可以达到节油不冒黑烟. 小化肥工厂搞优选, 既节煤又增产. 在大型化工设备上搞优选, 提高收率潜力更大. 解放牌汽车优选了化油器的合理尺寸, 一辆汽车一年可节油一吨左右, 全国现有民用汽车都来推广, 一年就可节油六十余万吨. 粮米加工优选加工工艺, 一般可提高出米率百分之一、二、三, 提高出粉率百分之一. 若按全国人数的口粮加工总数计算, 一年就等于增产几亿斤粮食.

最好的生产工艺是客观存在的, 优选法不过是提供了认识它的、尽量少做实验、快速达到目的的一种数学方法.

物资的合理调配, 农作物的合理分布, 水库的合理排灌, 电网的合理安排, 工业的合理布局, 都要用到数学才能完满解决, 求得合理的方案. 总之一句话, 在具有各种互相制约、互相影响的因素的统一体中, 寻求一个最合理 (依某一目的, 如最经济, 最省人力) 的解答便是一个数学问题, 这就是"多、快、好、省"原则的具体体现. 所用到的数学方法很多, 其中确属适用者我们也准备了一些, 但由于林彪、"四人帮"一伙的干扰破坏, 没有力量进行深入的工作. 今天, 在开创社会主义建设事业新局面的同时, 数学研究和应用也必将出现一个崭新的局面.

## 数学之发展

宇宙之大, 粒子之微, 火箭之速, 化工之巧, 地球之变, 生物之谜, 日用之繁, 无处不用数学. 其他如爱因斯坦用了数学工具所获得的公式指出了寻找新能源的方向, 并且还预示出原子核破裂发生的能量的大小. 连较抽象的纤维丛也应用到了物理当中. 在天文学上, 也是先从计算上指出海王星的存在, 而后发现了海王星. 又如高速飞行中, 由次音速到超音速时出现了突变, 而数学上出现了混合型偏微分方程的研究. 还有无线电电子学与计算技术同信息论的关系, 自动化与控制技术同常微分方程的关系, 神经系统同控制论的关系, 形态发生学与结构稳定性的关系等等不胜枚举.

数学是一门富有概括性的学问. 抽象是它的特色. 同是一个方程, 弹性力学上是描写振动的, 流体力学上却描写了流体动态, 声学家不妨称它是声学方程, 电学家也不妨称它为电报方程, 而数学家所研究的对象正是这些现象的共性的一面 —— 双曲型偏微分方程. 这个偏微分方程的解答的性质就是这些不同对象的共同性质, 数值的解答也将是它所联系各学科中所要求的数据.

不但如此, 这样的共性, 一方面可以促成不同分支产出统一理论的可能性, 另一方面也可以促成不同现象间的相互模拟性. 例如: 声学家可以用相似的电路来研究声学现象, 这大大地简化了声学实验的繁重性. 这种模拟性的最普遍的应用便是模拟电子计算机的产生. 根据神经细胞有兴奋与抑制两态, 电学中有带电与不带电两态, 数学中二进位数的 0 与 1、逻辑中的"是"与"否", 因而有用电子数字计算机来模拟神经系统的尝试, 及模拟逻辑思维的初步成果.

我们作如上的说明, 并不意味着数学家可以自我陶醉于共性的研究之中. 一方面我们得承认, 要求数学家深入到研究对象所联系的一切方面是十分困难的, 但是这并不排斥数学家应当深入到他所联系到的为数众多的科学之一或其中的一部分. 这样的深入是完全必要的. 这样做既对国民经济建设可以做出应有的贡献, 而且就是对数学本身的发展也有莫大好处.

客观事物的出现一般讲来有两大类现象. 一类是必然的现象 —— 或称因果律.

一类是大数现象 —— 或称机遇律. 表示必然现象的数学工具一般是方程式, 它可以从已知数据推出未知数据来, 从已知现象的性质推出未知现象的性质来. 通常出现的有代数方程, 微分方程, 积分方程, 差分方程等等 (特别是微分方程). 处理大数现象的数学工具是概率论与数理统计. 通过这样的分析便可以看出大势所趋, 各种情况出现的比例规律.

数学的其他分支当然也可以直接与实际问题相联系. 例如: 数理逻辑与计算机自动机的设计, 复变函数论与流体力学, 泛函分析与群表示论之与量子力学, 黎曼几何之与相对论等等. 在计算机设计中也用到数论. 一般说来, 数学本身是一个互相联系的有机整体, 而上面所提到的两方面是与其他科学接触最多, 最广泛的.

计算数学是一门与数学的开始而俱生的学问, 不过今天由于快速大型计算机的出现特别显示出它的重要性. 因为对象日繁, 牵涉日广 (一个问题的计算工作量大到了前所未有的程度). 解一个一百个未知数的联立方程是今天科学中常见的 (如水坝应力, 大地测量, 设计吊桥, 大型建筑等等), 仅靠笔算就很困难. 算一个天气方程, 希望从今天的天气数据推出明天的天气数据, 单凭笔算要花成年累月的时间. 这样算法与明天的天气何干? 一个讽刺而已! 电子计算机的发明就满足了这种要求. 高速度大存储量的计算机的发展改变了科学研究的面貌, 但是近代的电子计算机的出现丝毫没有减弱数学的重要性, 相反地更发挥数学的威力, 对数学的要求提得更高. 繁重的计算劳动减轻了或解除了, 而创造性的劳动更多了. 计算数学是一个桥梁, 它把数学的创造同实际结合起来. 同时它本身也是一个创造性的学科. 例如推动了一个新学科计算物理学的发展.

除掉上面所特别强调的分支以外, 并不是说数学的其余部分就不重要了. 只有这些重点部分与其他部分环环扣紧, 把纯数学和应用数学都分工合作地发展起来, 才能既符合我国当前的需要, 又符合长远需要.

从历史上数学的发展的情况来看, 社会愈进步, 应用数学的范围也就会愈大, 所应用的数学也就愈精密, 应用数学的人也就愈多. 在日出而作, 日入而息的古代社会里, 会数数就可以满足客观的需要了. 后来由于要定四时, 测田亩, 于是需要窥天测地的几何学. 商业发展, 计算日繁, 便出现了代数学. 要描绘动态, 研究关系的变化, 变化的关系, 因而出现了解析几何学、微积分等等.

数学的用处在物理科学上已经经过历史考验而证明. 它在生物科学和社会科学上的作用也已经露出苗头. 存在着十分宽广的前途.

最后, 我得声明一句, 我并不是说其他科学不重要或次重要. 应当强调的是, 数学之所以重要正是因为其他科学的重要而重要的, 不通过其他学科, 数学的力量无法显示, 更无重要之可言了.

需要指出的是, "四人帮" 为了复辟资本主义, 疯狂地破坏生产, 破坏科学技术的发展, 他们既破坏理论研究工作, 更疯狂地打击从事应用数学的工作者. 他们的

遗毒需要彻底清除, 不可低估. 为了实现"四个现代化", 把我国建成强大的社会主义国家这一伟大目标, 发展数学的重要性是无可置辩的.

# 有限与无穷, 离散与连续①

## —— 为纪念中国科学技术大学建校五周年而作

### 一

这是我们教低年级数学基础课的一些体会, 似乎是看出了些问题, 但由于作者的水平限制, 对数学的了解是片面的, 并且更没有哲学修养能从若干感性知识中概括出理性论断来. 所以写这样一篇提供素材的文章, 希望聚沙成塔, 集腋成裘, 以备沙里淘金者的参考.

数学中有两大类的问题: 一类是离散性质的, 一类是连续性质的. 在我们一生学习的过程中, 开始于数数 —— 一、二、三、四、五、······. 这完全是离散性质的东西. 算术、代数都是处理离散性质问题的学科. 整个中学阶段所学的教学可以说都不是突出利用"连续性"与"无穷性"的学科. 直线上的点显示出连续性质, 但突出地重用"连续"与"无穷"确始于微积分. 在描绘一瞬间的速度, 或一瞬间的量的变化, 我们重用了"连续性". 这就是初等数学与高等数学的分界. 但如果从"初等"、"高等"这些字样, 或我们学习的次序, 就断定"连续性的数学"比"离散性的数学"更优越了或更能解决问题了, 那就不尽然了. 本文的目的在于着重地谈谈离散性的重要. 但必须指出, 我们不是说连续性次重要些, 而是说必须两者妥善结合. 一切从实际出发, 看需要而决定. 不能强调一面而忽略一面, 但有一点似乎可以向初学者建议的, 在学连续性数学之前, 先打好所对应的离散性数学的基础. 因为绝大部分连续性的结果往往以离散性的结果做背景的, 或者是离散性问题的极限. 但并不是说, 我们不应当把学习的时间或精力在连续性数学上多花一些.

先看看客观事实, 如果本来就是离散的, 那就不必人为地引进连续性 (但并不排斥, 虽然离散, 但多到无法处理的时候, 也势所必致地用连续方法来处理的可能性, 如沙的流动). 在资本主义国家里面有些经济学者, 用微分方程来处理经济学上的问题, 我们对经济学一窍不通, 不能有所批判, 但有一点可以肯定, 他们所根据的数据是离散 —— 或者实质上不可能连续化的. 如: 农业生产量不能分为每瞬间几何? 它是季度性生产, 连分月份都不可能, 枉论其他. 用连续方法来处理离散问题, 对头否? 但他们有这样的答辩: 用上了微分方程就有定性理论, 利用它易于看出发

① 本篇与王元同志合作. 感谢中国科学院裴丽生副院长的鼓励. 他建议我们把教学体会不要仅仅写在数学著作 [1] 或教材 [2] 中, 把一些与其他兄弟学科可能有关的东西, 写出来互相交流, 因此才写了这样一篇内容芜杂的文章, 敬求兄弟学科及本学科同志们的指教.

展趋势. 岂其然哉! 实质上, 利用差分方程或矩阵乘方的性质照样可以看到趋势. 并且还容易些, 还浅显些. 但是在大学课程中没有包括进去而已, 或原则上有之, 但未像微分方程那样多方强调而已. 在 (三) 中还将指出连续化的不可能性, 硬用较深的数学殊无谓也. 深入浅出是功夫, 浅入深出是浪费.

我们有这样的不成熟的看法, 先学些矩阵知识, 差分方程, 再学微分方程, 则既可以学得处理"离散"问题的方法, 取其极限, 往往又可以得出微分方程的结果.

以上所讲就是说: 离散问题用离散方法来处理为妥的论点. 现在进一步说明: 连续问题中的离散处理方法.

首先的问题是数据取得的问题. 能不能取得无穷精密的数据? 不能, 即使准到十位百位, 用十位百位小数表达出来的数据所成的集体仍然是离散的, 而不是连续的 (并且有时过分的精密度是完全不必要的). 再则取数据的次数也必然是有限的, 离散的.

其次看计算工具, 近代的数字电子计算机本质上是离散的. 它的特点是根据有限位数据进行有限次运算, 算出有限个有限位的解答来. 一切有限, 仍然是离散的.

最后所能拿出来的结果 (或客观的要求也是如此) 当然也是离散的. 这是一个从离散到离散的过程. 数学家们通常的想法是从离散数据用插入法或回归法得函数, 得微分方程, 微分方程直接解不出来, 再将微分方程差分化变为代数方程 (离散), 然后得出离散性的解答来. 其过程中, 经过插入法有误差, 经过差分法又有误差, 变成代数问题以后的求解误差就不提了. 因而提出了以下的课题: 能不能从离散直接到离散. 这样避免了经过函数逼近的误差, 避免了经过微分方程差分求解的误差. 如果可能, 则方法初等化了! 而结果反而可能更精密了! 我们水平限制不敢多所论列, 但主观上谬认为这是一个值得尝试的方向. 再申明一下, 重视离散性方法的同时, 我们绝不能忽视连续性方法. 解析数论就是一门用连续性方法处理离散问题而获得重要成果的分支. 连续性的考虑往往会看到一些离散性所不易看到的问题.

以下罗列一些例子, 这些例子是从教基础课得来的. 选择的标准当然也就是基础课或略高一些的水平. 并且都是选取了与其他学科的科学工作者有共同兴趣的问题. 各节之间的关系也是不太大的. 例如: 常用傅里叶级数的同志不妨看看第四节.

再重复一句, 这是抛砖引玉性质的文章. 多举出些具体的感性材料, 有可能为将来的教学改革或理论认识创造条件. 虚心求教, 敬请指正.

## 二　对象是连续的, 但我们只能了解到其有限个数据 —— 算体积, 算面积

在学了微积分之后, 我们常常有这样的喜悦: 任何曲线的长度, 任何曲面的面

积及任何物体的体积都可以用积分方法来处理了. 这种喜悦是应当有的, 也是可以理解的. 但是以为这就已经可以解决问题了, 那就错了. 深入一想, 我们所学过的方法都有一个共同的要求, 就是要求有表示曲线、曲面的公式. 也就是在实际中, 有没有这样的表达公式? 例如说, 在估计矿藏储量时, 有没有一个表示这矿体周界的解析公式. 又如在估计山坡面积时, 有没有一个 $z=f(x,y)$ 表示这曲面的公式. 在实际情况中是没有的. 一来由于我们不可能对每一点都进行实测, 二来由于即使对矿体测了很多点, 但也是不能够求出曲面的表达式来的, 即使拼拼凑凑找出个公式, 但在求积分的时候, 依然是积不出来 (找原函数) 的时候多, 而能够积成初等函数的时候少 —— 少得很. 因而矿体和山坡虽然是连续分布的, 但是我们还是必须用离散的方法才能 (近似) 估出体积及面积.

但这并不是说微积分里求面积体积的公式没有用了, 这儿是说, 必须看看怎样才能用得上, 并且将发现, 理论是有用的, 它能给我们提供具体的线索, 并帮我们判断各种方法的优劣性及进一步改善这些方法.

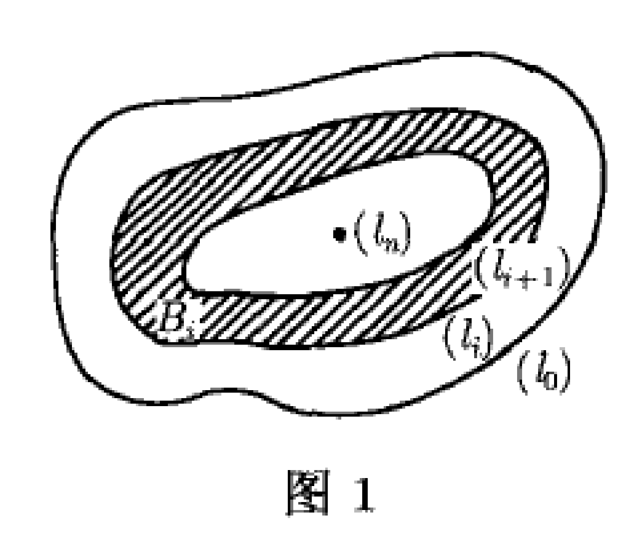



图 1

还是举一个例子吧: 在估计山坡面积时, 有两套方法: 一套是地理学上的方法, 称为 Волков法, 另一套是矿藏几何学上的方法, 称为Бауман法. 以下我们把它们介绍一下, 再比优劣.

假定地图上以 $\Delta h$ 为高程差画出等高线, 并假定有一制高点及等高线成圈 (其他情况很容易由此被推导出来). 假定由制高点 $(l_n)$ 向外一圈一圈地画等高线 $(l_{n-1})$, $(l_{n-2})$, $\cdots$, $(l_0)$. 取 $(l_0)$ 的高度为 0, $(l_n)$ 的高度为 $h$, $(l_i)$ 与 $(l_{i+1})$ 之间的面积用 $B_i$ 表示 (即投影的面积).

1. Бауман方法.

a) $C_i=\dfrac{1}{2}(l_i+l_{i+1})\Delta h$(中间直立隔板的面积)[①];

b) $\displaystyle\sum_{i=0}^{n-1}\sqrt{B_i^2+C_i^2}$ 就是所求的斜面积的近似值.

2. Болков 方法.

① 等高线 $(l_i)$ 的长度用 $l_i$ 表示.

a) $l=\sum_{i=0}^{n-1} l_i$ 为等高线的总长度. $B=\sum_{i=0}^{n-1} B_i$ 为总投影面积. 由

$$\operatorname{tg} \alpha=\frac{\Delta h \cdot l}{B}$$

得出平均倾角 $\alpha$;

b) $B \sec \alpha=\sqrt{B^2+(\Delta h \cdot l)^2}$ 就是所求的斜面积的近似值.

这两个方法哪一个更好一些? 这些方法所给出的结果在怎样的程度上迫近斜面积? 又当等高线的分布趋向无限精密时, 这些方法所给出的结果是什么? 是否就是真的面积? 下面我们将回答这些问题.

以制高点为中心引进极坐标. 命高度是 $z$ 的等高线方程是

$$\rho=\rho(z, \theta), \ 0 \leqslant \theta \leqslant 2 \pi$$

(假定 $\rho(z, \theta)$ 适当地光滑). 命 $z_i=\frac{h}{n} i, \ \Delta h=\frac{h}{n}$. 则 $(l_i)$ 所围绕的面积等于

$$\frac{1}{2} \int_0^{2 \pi} \rho^2(z_i, \theta) d \theta .$$

$(l_i)$ 的长度等于

$$l_i=\int_0^{2 \pi} \sqrt{\rho^2(z_i, \theta)+\left(\frac{\partial \rho(z_i, \theta)}{\partial \theta}\right)^2} d \theta .$$

于是由中值公式得

$$B_i=-\int_0^{2 \pi} \rho(z_i', \theta) \frac{\partial \rho(z_i', \theta)}{\partial z_j'} d \theta \ \Delta h$$

及

$$C_i=\int_0^{2 \pi} \sqrt{\rho^2(z_i'', \theta)+\left(\frac{2 \rho(z_i'', \theta)}{\partial \theta}\right)^2} d \theta \ \Delta h,$$

其中 $z_i \leqslant z_i', \ z_i'' \leqslant z_{i+1}$. 因此当 $\Delta h \rightarrow 0$ 时, $\sum_{i=0}^{n-1} \sqrt{B_i^2+C_i^2}$ 趋近于

$$\text{Бa}=\int_0^h \sqrt{\left(\int_0^{2 \pi} \rho \frac{\partial \rho}{\partial z} d \theta\right)^2+\left(\int_0^{2 \pi} \sqrt{\rho^2+\left(\frac{\partial \rho}{\partial \theta}\right)^2} d \theta\right)^2} d z .$$

这便是当 $\Delta h \rightarrow 0$ 时, 用Бауман方法算出的斜面积所趋近的值. 而 $\sqrt{\left(\sum_{i=0}^{n-1} B_i\right)^2+\left(\Delta h \sum_{i=0}^{n-1} l_i\right)^2}$ 的极限

$$\mathrm{B}_0=\sqrt{\left(\int_0^{2 \pi} d \theta \int_0^h \rho \frac{\partial \rho}{\partial z} d z\right)^2+\left(\int_0^h d z \int_0^{2 \pi} \sqrt{\rho^2+\left(\frac{\partial \rho}{\partial \theta}\right)^2} d \theta\right)^2}$$

便是 Волков方法算出的斜面积所趋近的值.

习知曲面的面积 $S$ 为

$$S=\int_0^h\int_0^{2\pi}\sqrt{\rho^2+\left(\frac{\partial\rho}{\partial\theta}\right)^2+\left(\rho\frac{\partial\rho}{\partial z}\right)^2}d\theta\ dz.$$

引入一个复值函数

$$f(z,\theta)=-\rho\frac{\partial\rho}{\partial z}+i\sqrt{\rho^2+\left(\frac{\partial\rho}{\partial\theta}\right)^2},$$

则

$$\begin{aligned}S&=\int_0^h\int_0^{2\pi}|f(z,\theta)|d\theta\ dz,\\ \text{Бa}&=\int_0^h\left|\int_0^{2\pi}f(z,\theta)d\theta\right|dz,\\ \text{B}_0&=\left|\int_0^h\int_0^{2\pi}f(z,\theta)d\theta\ dz\right|.\end{aligned}$$

由此可见:

$$\text{B}_0\leqslant\text{Бa}\leqslant S.$$

**结论** (i) Бауман方法比 Волков 方法精密些, (ii) 所求出的结果比真正的结果常常偏低一些.

除此而外, 不难讨论 $\text{B}_0=S$ 及$\text{Бa}=S$ 的情况. 我们还可以给出由这些方法所产生的误差的估计, 并指出产生误差的原因及避免误差的方法. 关于这些请参看 [1], [2].

**附记 1** 本节所用的积分是可以避免的.

## 三 无法连续化 —— 非负方阵

如产量, 如能量, 如概率都不能是负数. 在宇宙线的簇射过程中, 在运筹学及概率论的若干问题中, 往往出现非负元素的方阵, 即某些物态的多寡经过某段时间之后的变化情况可以用非负方阵表达之. 更具体些说, 例如有甲乙丙三种物件各有 $a,b,c$ 单位. 但是经过一段时间 $t$ 之后, 甲类物质变为 $ap_{11},ap_{12},ap_{13}$ 单位的甲乙丙三类物质, 而乙类物质变为 $bp_{21},bp_{22},bp_{23}$ 单位的甲乙丙三类物质; 丙类物质变为 $cp_{31},cp_{32},cp_{33}$ 单位的甲乙丙三类物质. 即经过时间 $t$ 后, 甲乙丙物质的数量各为

$$ap_{11}+bp_{21}+cp_{31},$$

$$ap_{12}+bp_{22}+cp_{32},$$

$$ap_{13}+bp_{23}+cp_{33}$$

个单位. 由于物质不能变负, 所以 $p_{ij}\geqslant 0$. 这方阵

$$P=\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

称为变化方阵. 如果原始物质的数量用矢量 $v=(a,b,c)$ 表之, 则经过时间 $t$ 后, 其数量将为

$$vP.$$

如果仍然照这样的关系变化, 则经过 $2t$ 时间将得

$$vP^2.$$

经过 $nt$ 时间则为

$$vP^n.$$

当 $n$ 增加时, 我们可以看出发展趋势.

为了易于了解起见, 我们回到单一的情况. 设原来的数量是 $c$, 经过单位时间后变为 $cq$, 经过 $n$ 个单位时间得 $cq^n$, 经过半个单位时间可以设想, 它的数量是 $cq^{1/2}$(注意问题就在这儿了！), 一般地讲, 可以设想在时间 $t$ 的时候, 它的数量是 $f(t)=cq^t$, 它的微分表达式是

$$\frac{df}{dt}=(\log q)f, f(0)=c.$$

也就是说 $f(t)=cq^t$ 是微分方程唯一的解.

对于单一的现象, 这方法虽有在理论上不妥当的地方, 即在时间 $1/2$ 是否是 $q^{1/2}$ 倍. 但是在应用的时候并不出现困难, 其主要原因是一个正数可以任意开方, 也就是

$$\lim_{\varepsilon\to 0}\frac{q^{\varepsilon}-1}{\varepsilon}=\log q$$

是实的存在的. 这个规律可以描述为**量的增加率与时间成比例**. 因此可能事实上虽然 $q^{1/2}$ 不定义, 但我们理想地设想它存在, 并不会发生什么矛盾.

如果有人希望把这一规律推广到多个现象的时候, 那就势所必然地要求, 求方阵 $P$ 的平方根; 求方阵 $P$ 的任意方根. 是否有非负方阵的平方等于 $P$? 如果没有, 则用微分处理是不可能的, 举个例子, 没有非负方阵的平方等于

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

其理由是极简单的，如果

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

则 $a^2+bc=0$，由 $a,b,c$ 的非负性质得 $a=0$ 及 $b$ 或 $c=0$. 这是不可能的，换言之，如果方阵 $P$ 不是“无穷可分的”，也就是没有实方阵 $Q$ 使 $P=e^Q$，则不可能用微分方法来处理. 在研究线性弹性系统微振动的颤动性质的时候，所对应的方阵的特征根全是正的，而且是不同的，因而 $Q$ 是存在的. 但在经济现象中，有波浪式前进，螺旋式上升的现象，这说明它所对应的方阵不可能全部是正根，而可能有负根或复根存在，如果出现负根即就无法保证“无穷可分”性. 因而用微分方程的理论来笼统地处理经济现象是欲巧反拙的.

在物理现象及概率现象中，当运用微分方程来处理这种现象的时候，既要考虑能不能，又要考虑要不要，如果并不能证明“无穷可分”时，用差分方程保险些. 在证明了“无穷可分”时，也可能用差分方程更简单些. 不一定要用微分方程.

这儿再说些题外之言，完成演变所需要的时间是否有“单位”存在？即短于这个时间，不能完成某种演变. 在这样的情况下，“差分”法比“微分”法更能表达客观现象. 在这种现象中，时间变为“离散”. 但基本单位是多长？如果多种不同单位现象的混合，情况又如何？在数学上反映出来更有可度约与不可度约的情况. 因而类似数论中 Diophantine 逼近的现象出现了，但确是远更复杂的问题.

**附记 1** 关于非负方阵的一些性质.

**定理 1** 如果

$$\sum_{i=1}^{n} a_{ij} \leqslant q,\ a_{ij} \geqslant 0,\ j=0,\ 2,\ \cdots,\ n, \tag{1}$$

则方阵 $A=(a_{ij})$ 的特征根的绝对值都 $\leqslant q$.

这定理的证明是很简单的. 由特征根 $\lambda$ 的定义，有非全为零的 $x_1,\cdots,x_n$ 使

$$\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i.$$

因此

$$|\lambda||x_i| \leqslant \sum_{j=1}^{n} a_{ij}|x_j|,$$

所以

$$|\lambda|\sum_{i=1}^{n}|x_i| \leqslant \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}\right)|x_j| \leqslant q\sum_{j=1}^{n}|x_j|,$$

即得所证.

这一性质在以后要用, 所以给予证明, 实质上, 非负方阵的若干性质与特征, 似乎都有它的经济学 (或其他用得到它的学科) 上的重要意义. 例如, 非负方阵有一个最大正特征根, 这似乎可以用来作为一个经济体系的发展速度的标志, 而对应于这特征根有一非负元素的特征矢量, 这特征矢量似乎反应了各种产品之间, 或产品与劳动之间的正确等价关系, 如果有复虚数的特征根存在, 则反映了可能若干部门间会出现螺旋式上升, 波浪式前进的情况.

不仅如此, 还可以提供 "应当改进哪些系数 (如每吨钢的煤耗系数) 可能使我们的经济系统增长最快" 的线索. 因而决定应当改进的关键性的环节. 当然这样的建议只能作为参考, 而更重要的是人的作用.

## 四　多算了反而吃亏 —— 实用调和分析

在广泛的应用中, 我们经常要把一个函数 $f(x)$ 展开成为 Fourier 级数, 即

$$f(x) \sim \frac{a_0}{2} + \sum_{m=1}^{\infty} (a_m \cos mx + b_m \sin mx), \tag{1}$$

这儿

$$a_m = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos mx \, dx,$$
$$b_m = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin mx \, dx. \tag{2}$$

有时用等价的复数形式的 Fourier 级数

$$f(x) \sim \sum_{m=-\infty}^{\infty} C_m e^{imx},$$
$$C_m = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-imx} dx. \tag{3}$$

如果 $f(x)$ 是由实验得来的, 有时仅测得有限个数据, 根据这有限个数据, 怎样求出渐近的 Fourier 级数来呢? 有时 $f(x)$ 即使有解析表达式, 但积分 (2), (3) 的原函数无法获得, 因而必须进行数值积分, 对于这两种情况, 一般都用以下的方法来处理.

假定在 $[0,2\pi]$ 中给了 $n(=2n'+1)$ 个点的函数值

$$y_l = f\left(\frac{2\pi l}{n}\right), \ 0 \leqslant l \leqslant n-1.$$

而用

$$a_m \sim a'_m = \frac{2}{n}\sum_{l=0}^{n-1} y_l \cos\frac{2\pi lm}{n}$$

及

$$b_m \sim b'_m = \frac{2}{n}\sum_{l=0}^{n-1} y_l \sin\frac{2\pi lm}{n}$$

来近似计算 $a_m$ 与 $b_m$. 也许会出现这样的错觉, 少取几个数据, 利用现代计算工具多算几项 $a'_m, b'_m$, 则

$$\frac{a_0}{2}+\sum_{m=1}^{N}(a'_m\cos mx+b'_m\sin mx) \tag{4}$$

会更精确地逼近于 $f(x)$. 这是不对的, 如果仅给了 $n$ 个数据, 即用 $n$ 个点的函数值来近似计算 $a_m$ 与 $b_m$. 过多的计算不但不能增加精确度, 反而会增大误差, 甚至于变成荒谬的结论. 其理由是 $a'_m = a'_{n+m}$, $b'_m = b'_{n+m} (m = 1, 2, \cdots)$, 所以级数

$$\frac{a'_0}{2}+\sum_{m=1}^{\infty}(a'_m\cos mx+b'_m\sin mx) \tag{5}$$

是发散的. 因而一直算下去, 所得出的结果将大大偏离于原来所给的函数 $f(x)$(特别当 $f(x)$ 是有一定光滑的函数, 例如连续, 可微商等). 我们可以证明最好是算到 $n$ 项, 多算则浪费精力, 造成更大的误差, 少算则没有充分利用数据.

用初等指数和的方法来处理这一问题, 方法是离散性的, 并且亦易于计算, 先从复数形式的 Fourier 级数演算起: 假定在区间 $[-\pi,\pi]$ 中给了函数 $f(x)$ 的 $n(=2n'+1)$ 个数据

$$y_l = f\left(\frac{2\pi l}{n}\right),\ l=0,\ \pm1,\ \cdots,\ \pm n'. \tag{6}$$

利用公式

$$\frac{1}{n}\sum_{l=-n'}^{n'} e^{2\pi ilm/n} = \begin{cases} 0, & 若 n\nmid m, \\ 1, & 若 n|m. \end{cases} \tag{7}$$

可以从

$$y_l = \sum_{m=-n'}^{n'} C'_m e^{2\pi ilm/n}, |l|\leqslant n' \tag{8}$$

定出 $C'_m$ 来. 定 $C'_m$ 的方法是: 以 $e^{-2\pi ilq/n}$ 乘 (8) 式, 并对 $l$ 求和, 由 (7) 得出

$$\sum_{l=-n'}^{n'} y_l\, e^{-2\pi ilq/n} = \sum_{m=-n'}^{n'} C'_m \sum_{l=-n'}^{n'} e^{2\pi i(m-q)l/n} = n\, C'_q. \tag{9}$$

因此建议我们用

$$S_n(x)=\sum_{n=-n'}^{n'}C'_m e^{imx},$$

$$C'_m=\frac{1}{n}\sum_{l=-n'}^{n'}y_l e^{-2\pi ilm/n} \tag{10}$$

来逼近 $f(x)$. 我们现在来估计 $S_n(x)$ 与 $f(x)$ 的误差.

**定理 1** 假定 $f(x)$ 在 $[-\pi,\pi]$ 中有 $r(\geqslant 2)$ 阶连续微商, 而且是以 $2\pi$ 为周期的函数, 并且假定

$$|f^{(r)}(x)|<C,$$

则

$$|f(x)-S_n(x)|<\frac{4C}{(r-1)n'^{r-1}}. \tag{11}$$

**证** 已知

$$f(x)=\sum_{m=-\infty}^{\infty}C_m e^{imx},$$

$$C_m=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(x)e^{-imx}dx. \tag{12}$$

分部积分 $r$ 次得

$$C_m=\frac{1}{2\pi(im)^r}\int_{-\pi}^{\pi}f^{(r)}(x)e^{-imx}dx.$$

立刻推得

$$|C_m|\leqslant\frac{C}{|m|^r}.$$

因此

$$\left|f(x)-\sum_{m=-n'}^{n'}C_m e^{imx}\right|\leqslant 2\sum_{m=n'+1}^{\infty}\frac{C}{|m|^r}\leqslant 2C\int_{n'}^{\infty}\frac{dx}{x^r}$$
$$=\frac{2C}{(r-1)n'^{r-1}}. \tag{13}$$

当 $|m|\leqslant n'$ 时,

$$C_m-C'_m=C_m-\frac{1}{n}\sum_{l=-n'}^{n'}y_l e^{-2\pi ilm/n}$$

$$=C_m - \frac{1}{n}\sum_{l=-n'}^{n'} e^{-2\pi ilm/n}\sum_{q=-\infty}^{\infty} C_q c^{2\pi iql/n}$$

$$=C_m - \frac{1}{n}\sum_{q=-\infty}^{\infty} C_q \sum_{l=-n'}^{n'} e^{2\pi i(q-m)l/n}$$

$$=C_m - \sum_{\substack{q=-\infty \\ q\equiv m (\mathrm{mod}\ n)}}^{\infty} C_q,$$

因此

$$|C_m - C'_m| \leqslant {\sum_{t=-\infty}^{\infty}}' |C_{m+nt}|,$$

这儿 $\Sigma'$ 表示和号中除去 $t=0$ 一项. 因此

$$\left|\sum_{m=-n'}^{n'}(C_m - C'_m)e^{imx}\right| \leqslant \sum_{m=-n'}^{n'}{\sum_{t=-\infty}^{\infty}}'|C_{m+nt}|$$

$$\leqslant \sum_{m=-n'}^{n'}\sum_{t=-\infty}^{\infty}\frac{C}{|m+nt|^r} \leqslant 2C\sum_{l=n'+1}^{\infty}\frac{1}{l^r} \leqslant \frac{2C}{(r-1)n'^{r-1}}. \quad (14)$$

(任一整数 $l$ 可以唯一地表成为 $nt+m(|m|\leqslant n')$ 的形式, 但 $t\neq 0$, 这表达除去 $|l|\leqslant n'$ 以外的所有整数, 故得所云.)

因此由 (12), (13), (14) 得

$$|f(x) - S_n(x)| < \frac{4C}{(r-1)n'r-1}.$$

在实际计算的时候, $S_n(x)$ 还可以表达得更简单些.

$$S_n(x) = \sum_{m=-n'}^{n'} C'_m e^{imx} = \frac{1}{n}\sum_{m=-n'}^{n'}\sum_{l=-n'}^{n'} y_l e^{-2i\pi lm/n} e^{imx}$$

$$=\frac{1}{n}\sum_{l=-n'}^{n'} y_l \sum_{m=-n'}^{n'} e^{i(x-2\pi l/n)m}$$

$$=\frac{1}{n}\sum_{l=-n'}^{n'} y_l \frac{\sin\left(n'+\frac{1}{2}\right)(x-2\pi l/n)}{\sin\frac{1}{2}(x-2\pi l/n)}$$

$$=\frac{1}{n}\sum_{l=-n'}^{n'} y_l \frac{\sin\left(\frac{1}{2}nx-\pi l\right)}{\sin\frac{1}{2}(x-2\pi l/n)}$$

$$=\frac{\sin\frac{1}{2}nx}{n}\sum_{l=-n'}^{n'}\frac{(-1)^l y_l}{\sin\frac{1}{2}(x-2\pi l/n)}. \tag{15}$$

**附记 1** 如果分点

$$0\leqslant x_1<\cdots<x_n<2\pi$$

不是均匀的, 则可以由联立方程

$$\begin{cases}\dfrac{a_0'}{2}+\displaystyle\sum_{m=1}^{n'}(a_m'\cos mx_i+b_m'\sin mx_i)=y_i,\\ \qquad\qquad (1\leqslant i\leqslant n)\\ \dfrac{a_0'}{2}+\displaystyle\sum_{m=1}^{n'}(a_m'\cos mx+b_m'\sin mx)=y(x).\end{cases}$$

消去 $a_0', a_m', b_m'$ 而得出 $y$ 与 $y_1,\cdots,y_n$ 的关系.

## 五 差分方法 —— 连续与离散间一座常用的桥梁

在微分方程的求解中, 我们常用差分方法, 这是一个应用十分广泛的方法. 简言之, 这一方法是将微分方程的求解问题化为代数方程 (即所谓差分方程) 的求解问题. 为了简单起见, 作为例子, 我们现在扼要地介绍一下用这一方法来处理 Laplace 方程的 Dirichlet 问题的过程. 在求解差分方程时, 我们将要谈到代数方法与 Monte Carlo 方法, 并作一些分析比较.

问题: 命 $G$ 是一个有光滑周界的有界的平面单联通区域. 在它的边界 $\Gamma$ 上给了一个连续函数 $f(x,y)$, 求连续函数 $u(x,y)$ 适合于

(i) 在 $G$ 内满足 Laplace 方程

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0. \tag{1}$$

(ii) 在 $\Gamma$ 上取已给函数 $f$ 的值.

关于 $u$ 的近似求法的步骤如次:

I. 网格化. 在平面上作与坐标轴平行的两族曲线

$$x=mh, y=nh,$$

这儿 $h$ 是某一正数, 而 $m,n$ 过所有的整数值. 这样的区域 $G$ 当然为一些以 $h$ 为边长的正方形所覆盖. 正方形的顶点称为整点. 与 $G$ 有公共点的正方形所成的区域以

$G^*$ 表之. $G^*$ 是一多边形. 命 $Q$ 是 $G^*$ 的边界上的整点, 假定 $\Gamma$ 与 $Q$ 的最近点是 $P$(如果有许多点有相同的距离, 则可取其中的任意一点), 我们定义 $f(Q)=f(P)$. 这样在 $G^*$ 的边界 $\Gamma^*$ 的整点上都有了函数值 $f(Q)$.

II. 差分化. 用

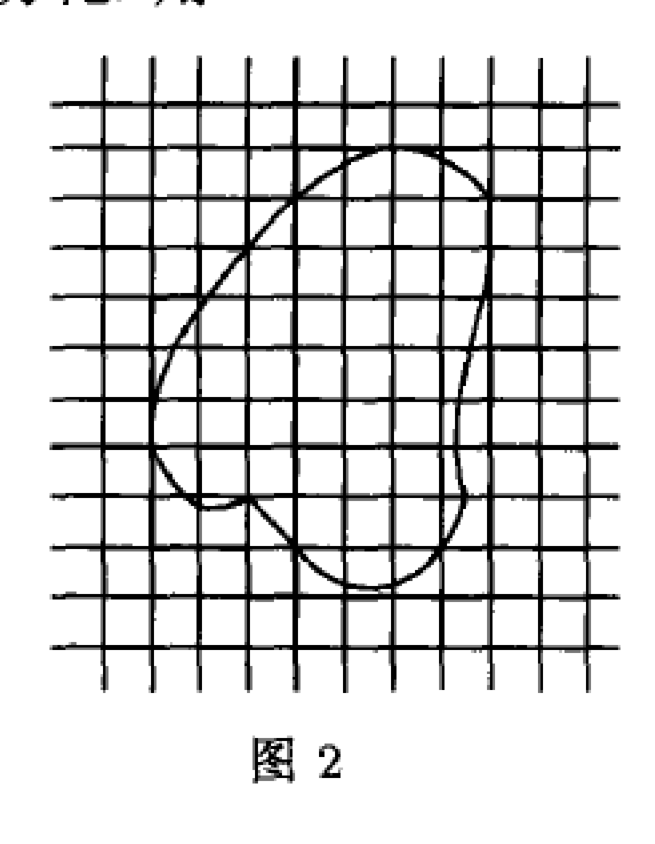


图 2

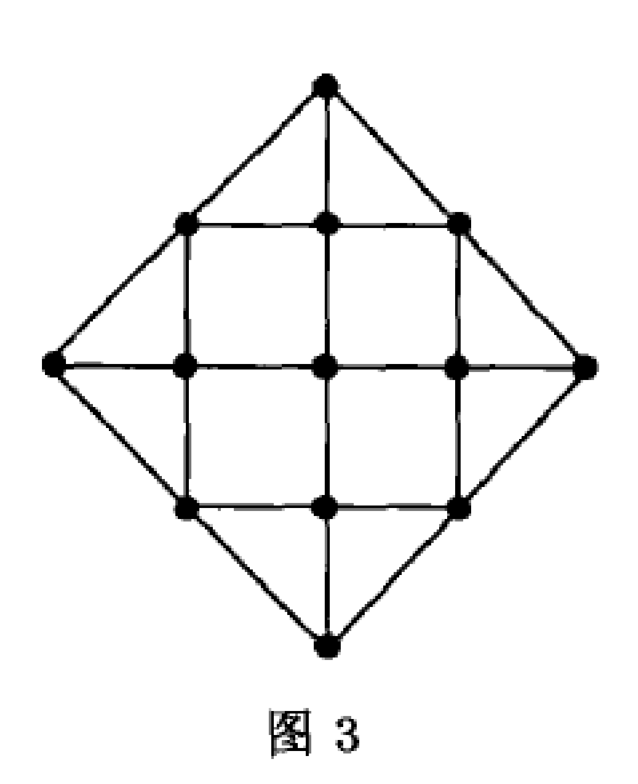


图 3

$$\frac{u(x+h,y)-2u(x,y)+u(x-h,y)}{h^2}$$

及

$$\frac{u(x,y+h)-2u(x,y)+u(x,y-h)}{h^2}$$

各代替二阶偏微商 $\dfrac{\partial^2 u}{\partial x^2}$ 及 $\dfrac{\partial^2 u}{\partial y^2}$, 则 Laplace 方程可以改写为

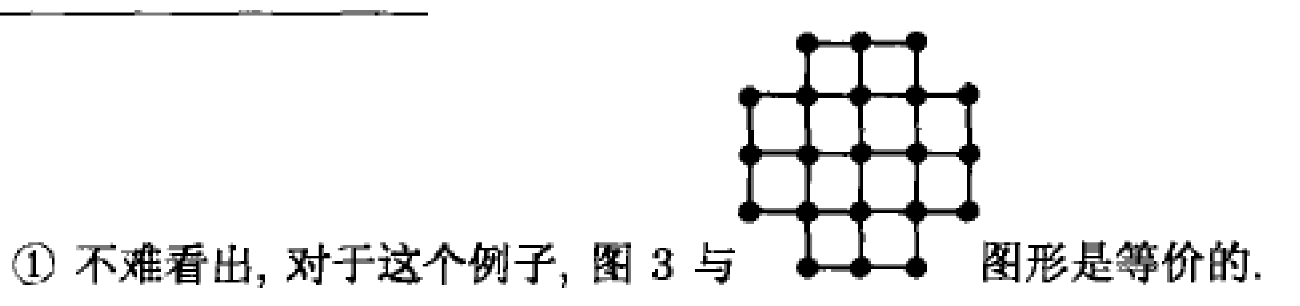
$$u(x,y)=\frac{1}{4}[u(x+h,y)+u(x-h,y)+u(x,y+h)+u(x,y-h)]. \qquad (2)$$

也就是在非边界整点 $(x,y)$, 函数 $u(x,y)$ 的数值等于其东、南、西、北四邻近整点的函数值的平均.

III. 问题一变而为已知多边形边界整点的函数值而求内部整点的函数值的问题了, 即问题化为求解线性方程组 (2).

但是由此得到的是否会是矛盾方程组? 是否仅有唯一的解? 都是必须解答的问题. 我们现在先举一个简单的例子, 然后就直觉地看出一般的理论了.

不妨取 $h=1$, 给了八点的函数值 $u(2,0)$, $u(1,1)$, $u(0,2)$, $u(-1,1)$, $u(-2,0)$, $u(-1,-1)$, $u(0,-2)$, $u(1,-1)$, 求 $u(0,0)$, $u(1,0)$, $u(0,1)$, $u(-1,0)$, $u(0,-1)$ 五值①.

① 不难看出, 对于这个例子, 图 3 与  图形是等价的.

将方程式全部列出：

$$\left.\begin{aligned}u(0,0) =& \frac{1}{4}(u(1,0)+u(0,1)\\&+u(-1,0)+u(0,-1))\\u(1,0) =& \frac{1}{4}(u(2,0)+u(1,1)\\&+u(0,0)+u(1,-1))\\u(0,1) =& \frac{1}{4}(u(1,1)+u(0,2)\\&+u(-1,1)+u(0,0))\\u(-1,0) =& \frac{1}{4}(u(0,0)+u(-1,1)\\&+u(-2,0)+u(-1,-1))\\u(0,-1) =& \frac{1}{4}(u(1,-1)+u(0,0)\\&+u(-1,-1)+u(0,-2)).\end{aligned}\right\} \tag{3}$$

由消去法得出

$$\begin{aligned}u(0,0) =&\frac{1}{12}(u(2,0)+u(0,2)+u(-2,0)+u(0,-2))\\&+\frac{1}{6}((u(1,1)+u(-1,1)\\&+u(-1,-1)+u(1,-1)),\end{aligned} \tag{4}$$

$$\begin{aligned}u(1,0) =&\frac{13}{48}u(2,0)+\frac{7}{24}(u(1,1)+u(1,-1))\\&+\frac{1}{24}(u(-1,1)+u(-1,-1))\\&+\frac{1}{48}(u(0,2)+u(-2,0)+u(0,-2))\end{aligned} \tag{5}$$

等等.

这些系数 $\frac{1}{6},\frac{1}{12},\frac{13}{48},\cdots$ 的意义是什么？我们以后再交代. 先作以下的代数处理, 把这十三个点的函数值作为一个列矢量的元素, 则 (3) 式可以写成

$$
\begin{bmatrix} u(0,0) \\ u(1,0) \\ u(0,1) \\ u(-1,0) \\ u(0,-1) \\ u(1,1) \\ u(-1,1) \\ u(-1,-1) \\ u(1,-1) \\ u(2,0) \\ u(0,2) \\ u(-2,0) \\ u(0,-2) \end{bmatrix}
=
\left[\begin{array}{ccccc:cccccccc}
0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{4} & & & & & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 \\
\frac{1}{4} & & 0 & & & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\
\frac{1}{4} & & & & & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 \\
\frac{1}{4} & & & & & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & \frac{1}{4} \\
\hdashline
 & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & & & 0 & & \\
 & & & & & & & 1 & & & & & \\
 & & 0 & & & & & & 1 & & & & \\
 & & & & & & & & & 1 & & & \\
 & & & & & & 0 & & & & 1 & & \\
 & & & & & & & & & & & 1 & \\
 & & & & & & & & & & & & 1
\end{array}\right]
\begin{bmatrix} u(0,0) \\ u(1,0) \\ u(0,1) \\ u(-1,0) \\ u(0,-1) \\ u(1,1) \\ u(-1,1) \\ u(-1,-1) \\ u(1,-1) \\ u(2,0) \\ u(0,2) \\ u(-2,0) \\ u(0,-2) \end{bmatrix}
\tag{6}
$$

可以抽象得

$$
\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} P & Q \\ O & I \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}. \tag{7}
$$

这里 $v$ 是边界整点的函数值所成的列矢量, 而 $u$ 是“内部整点”的函数值所成的列矢量. 这种表达法对一般的问题都对. 它实质上表达了两件事: (i) 内部整点的函数值可以表为其东, 南, 西, 北四邻近整点的函数值的平均. (ii) 边界点仍然是边界点.

因为一个整点只能是不超过四个整点的邻近点, 所以方阵 $P$ 的每列元素之和皆 $\leqslant 1$. 现在来证明 $P^2$ 的每列元素之和皆 $< 1$. 盖若不然, 由于 $P$ 的元素只能取 0 与 $1/4$ 二值, 故 $P$ 必包有子方阵

$$
\begin{bmatrix}
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4}
\end{bmatrix}.
$$

但是因为不能有两个不同的整点具有同样的东, 南, 西, 北四邻近整点, 所以这是不

可能的. 因此 $P^2$ 的每列元素之和皆 $<1$. 因此由三节可知 $P^2$ 的特征根的绝对值皆 $<1$, 从而

$$\lim_{n\to\infty} P^n = 0.$$

而且

$$Q + PQ + P^2Q + \cdots$$

收敛 (收敛于 $(I-P)^{-1}Q$). 将 (7) 式连续迭代 $n$ 次得

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} P & Q \\ O & I \end{pmatrix}^n \begin{pmatrix} u \\ v \end{pmatrix}$$
$$= \begin{pmatrix} P^nQ + PQ + P^2Q + \cdots + P^{n-1}Q \\ O \qquad\qquad I \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

命 $n\to\infty$, 则得

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} OQ + PQ + P^2Q + \cdots \\ O \qquad\qquad I \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

因此

$$u = (Q + PQ + P^2Q + \cdots)v = (I + P + P^2 + \cdots)Qv. \tag{8}$$

这就是问题的解答. 也就是当给了边界整点的函数值 $v$, 可以由 (8) 算出内部整点的函数值 $u$ 来, 这建议了以下的算法.

(A) 代数法. 把 $u$ 写成列矢量 $(u_1,\cdots,u_l)'$, $v$ 写成 $(v_1,\cdots,v_k)'$, 如果内部整点①$u_i$ 与边界整点 $v_i$ 相邻, 则在 $Q$ 的 $(i,j)$ 位置记上 1/4, 否则记上 0. 如果内部整点 $u_i$ 与 $u_j$ 相邻, 则在 $P$ 的 $(i,j)$ 位置记上 1/4, 否则记上 0. 这样得出 $P$ 与 $Q$, 用以下的格式算出 (8) 来.

| | | |
|---|---|---|
| | $Qv$ | |
| $P$ | $PQv$ | $R_1 = Qv + PQv$ |
| $P^2$ | $P^2R_1$ | $R_2 = R_1 + P^2R_1$ |
| $P^4$ | $P^4R_2$ | $R_3 = R_2 + P^4R_2$ |
| $P^8$ | $P^8R_3$ | $R_4 = R_3 + P^8R_3$ |
| $\cdots$ | $\cdots$ | $\cdots$ |

用到我们的例子上, 由于

$$P^3 = \frac{1}{4}P,$$

① 内部整点与边界整点亦分别记之为 $u_1,\cdots,u_l$ 与 $v_1,\cdots,v_k$, 请勿混淆.

所以

$$
\begin{aligned}
&Q+PQ+P^2Q+\cdots\\
&= \left(I+(P+P^2)\left(1+\frac{1}{4}+\frac{1}{4^2}+\cdots\right)\right)Q\\
&= \left(I+\frac{4}{3}(P+P^2)\right)Q\\
&= \begin{bmatrix}
\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12}\\
\frac{7}{24} & \frac{1}{24} & \frac{1}{24} & \frac{7}{24} & \frac{13}{48} & \frac{1}{48} & \frac{1}{48} & \frac{1}{48}\\
\frac{7}{24} & \frac{7}{24} & \frac{1}{24} & \frac{1}{24} & \frac{1}{48} & \frac{13}{48} & \frac{1}{48} & \frac{1}{48}\\
\frac{1}{24} & \frac{7}{24} & \frac{7}{24} & \frac{1}{24} & \frac{1}{48} & \frac{1}{48} & \frac{13}{48} & \frac{1}{48}\\
\frac{1}{24} & \frac{1}{24} & \frac{7}{24} & \frac{7}{24} & \frac{1}{48} & \frac{1}{48} & \frac{1}{48} & \frac{13}{48}
\end{bmatrix}.
\end{aligned}
$$

必须指出, 这儿所介绍的计算程序比解方程组的普通程序更快速些.

现在再来看看 (4), (5) 中系数的几何意义, 看一下 (4) 式中的 $\frac{1}{6}$ 及 $\frac{1}{12}$ 可能会想到: 由 (0,0) 出发到 (0,2) 有一条直路, 到 (1,1) 有两条路 "⌈" 与 "⌋", 一共有 12 条路, 因而到 (0,2) 的可能性是 $\frac{1}{12}$, 而到 (1,1) 的可能性是 $\frac{2}{12}=\frac{1}{6}$ 等等.

这种讲法是有道理的, 但不易推广. 请看下面的说法: 从 (0,0) 到其东, 南, 西, 北各邻近点的可能性各占 $\frac{1}{4}$. 但这四点均非边界整点, 因而由 (0,0) 一步到达边界的可能性是零.

任何一内点到其四邻点的可能性都是 $\frac{1}{4}$. 因此从 (0,0) 走两步, 共 16 种可能性, 其中到一顶点的各有一种 (共四种), 到一边点的各有二种 (共八种). 进一步退一步仍在原点的四种. 因此任意走两步达到每一顶点的可能性是 $\frac{1}{16}$, 达到每一边点的可能性是 $\frac{2}{16}=\frac{1}{8}$, 仍回原地的可能是 $\frac{1}{4}$.

走三步不可能由 (0,0) 到达边界点.

再看走四步的情况, 走两步已达边界的情况不谈了. 后二步依然从 (0,0) 出发, 但现在到边界点的可能性要乘上 $\frac{1}{4}$ 了. 即由 (0,0) 走四步达到每一顶点的可能性是 $\frac{1}{4}\cdot\frac{1}{16}$, 达到每一边点的可能性是 $\frac{1}{4}\cdot\frac{1}{8}$. 而返回原地的可能性是 $\frac{1}{4}\cdot\frac{1}{4}$.

五步不能, 而六步的可能性各为

$$\frac{1}{4^2}\cdot\frac{1}{16},\ \frac{1}{4^2}\cdot\frac{1}{8},\ \frac{1}{4^2}\cdot\frac{1}{4}.$$

等等. 由 (0,0) 出发走奇数步达到边界点的可能性是没有的. 走 $2l$ 步达到每一顶点

的可能是

$$\frac{1}{4^{l-1}}\cdot\frac{1}{16}.$$

达到每一边点的可能性是

$$\frac{1}{4^{l-1}}\cdot\frac{1}{8}.$$

而返回原地的可能性是

$$\frac{1}{4^{l}}.$$

因此, 达到每一顶点的可能性是

$$\begin{aligned}&\frac{1}{16}+\frac{1}{4}\cdot\frac{1}{16}+\frac{1}{4^2}\cdot\frac{1}{16}\cdots\\ &=\frac{1}{16}\left(1+\frac{1}{4}+\frac{1}{4^2}+\cdots\right)=\frac{1}{16}\left(1-\frac{1}{4}\right)^{-1}=\frac{1}{12}.\end{aligned}$$

而达到每一边点的可能性是

$$\frac{1}{8}\left(1+\frac{1}{4}+\frac{1}{4^2}+\cdots\right)=\frac{1}{6}.$$

返回原地的可能性是

$$\lim_{l\to\infty}\frac{1}{4^l}=0.$$

这就是概率论中的随机游动.

再看 (5) 式中 $\frac{13}{48}$ 的意义：由 (1,0) 一步可能到达 (2,0), 可能性是 $\frac{1}{4}$. 由 (1,0) 走一步不到边界点只可能到 (0,0), 可能性是 $\frac{1}{4}$, 以后的情况与从 (0,0) 出发相同. 因此走一步以上达到 (2,0) 的可能性是 $\frac{1}{4}\cdot\frac{1}{12}$. 总的可能性是

$$\frac{1}{4}+\frac{1}{4}\cdot\frac{1}{12}=\frac{13}{48}.$$

同样到 (1,1)(或 $(1,-1)$) 的可能性是

$$\frac{1}{4}+\frac{1}{4}\cdot\frac{1}{6}=\frac{7}{24}.$$

到其他的边界点必经 (0,0), 因此就是 (0,0) 到达这些点的可能性乘以 $\frac{1}{4}$, 即得

$$\frac{1}{4}\cdot\frac{1}{12}=\frac{1}{48}\text{ 与 }\frac{1}{4}\cdot\frac{1}{6}=\frac{1}{24}.$$

从这一简单的例子, 不难直觉地看出一般的理论. 这也建议我们用概率方法来解决"Laplace 方程的边界值问题". 实质上, 解决"差分化后的代数方程组".

(B) Monte Carlo 法 (或概率法). 我们先一般地定义二维随机游动如下：设有一质点从 $G^*$ 的某一整点出发, 以等概率 1/4 向其东, 南, 西, 北四相邻整点移动一步, 然后再以同样的方式, 从新的位置向其相邻的四整点处移动一步, 如此继续下去,

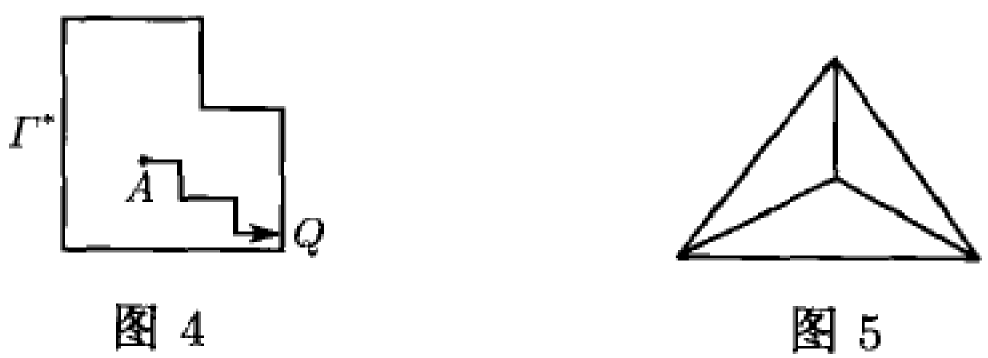



图 4　　　　图 5

直到达到某一边界整点, 游动便告终止. 设随机游动的一条路线是

$$\gamma_A : A \to A_1 \to \cdots \to A_{l-1} \to Q \in \Gamma^*,$$

则定义随机变量的值为

$$\xi = \xi(\gamma_A) = f(Q)$$

此处 $f(Q)$ 为边界整点 $Q$ 的函数值, 若 $Q$ 为边界整点, 则定义

$$\xi = \xi(\gamma_Q) = f(Q).$$

随机变量 $\xi$ 的数学期望 $E(\xi)$ 即方程组 (2) 的解. 换言之

$$E(\xi(\gamma_A)) = \frac{1}{4}\sum_{i=1}^{A} E(\xi(\gamma_{A_{1i}})),$$

$$\text{若} \quad A \in G^* \tag{9}$$

及

$$E(\xi(\gamma_Q)) = f(Q), \quad \text{若} \quad B \in \Gamma^*, \tag{10}$$

此处 $A_{11}$, $A_{12}$, $A_{13}$, $A_{14}$ 分别为 $A$ 的东, 南, 西, 北四邻点.

命 $P(\gamma_A)$ 表示循路线 $\gamma_A$ 游动的概率. 则

$$P(\gamma_A) = \frac{1}{4^l}.$$

因此

$$E(\xi(\gamma_A)) = \sum_{\gamma_A} \xi(\gamma_A) P(\gamma_A),$$

此处右端为对一切从 $A$ 出发的游动路线求和. 由 $A$ 出发, 第一步必然是走到其东, 南, 西, 北四邻点 $A_{11}$, $A_{12}$, $A_{13}$, $A_{14}$ 中的一个, 然后再继续游动. 因此

$$E(\xi(\gamma_A)) = \sum_{i=1}^{A} \sum_{\gamma_{A_{1i}}} \xi(\gamma_{A_{1i}}) P(A \to A_{1i}) P(\gamma_{A_{1i}}).$$

由于 $P(A \to A_{1i}) = \frac{1}{4}$, 所以

$$
\begin{aligned}
E(\xi(\gamma_A)) &= \frac{1}{4}\sum_{i=1}^{A}\sum_{\gamma_{A_{1i}}}\xi(\gamma_{A_{1i}})P(\gamma_{A_{1i}}) \\
&= \frac{1}{4}\sum_{i=1}^{A}E(\xi(\gamma_{A_{1i}})).
\end{aligned}
$$

此即 (9) 式. 其次当 $Q \in \Gamma^*$ 时, 只有一条游动路线, 即停止不动. 因此

$$E(\xi(\gamma_q)) = \sum_{\gamma_Q}\xi(\gamma_Q)P(\gamma_Q) = \xi(\gamma_Q) = f(Q).$$

故得 (10) 式.

设对 $\xi$ 进行了 $N$ 次观察得到

$$\xi_1, \cdots, \xi_N.$$

则根据大数定律可知, 对于任意 $\varepsilon > 0$ 皆有

$$\lim_{N\to\infty} P\left(\left|E(\xi) - \frac{1}{N}\sum_{i=1}^{N}\xi_i\right| \leqslant \varepsilon\right) = 1.$$

因此当 $N$ 充分大时

$$\frac{1}{N}\sum_{i=1}^{N}\xi_i$$

就可以作为 $E(\xi)$(即解答) 的近似值.

随机游动一般是用物理方法或者用数学方法产生的随机数来实现的. 在此不详谈了. 这里说一个通俗的办法: 用粉笔将 $G^*$ 画在围棋盘上. 如果要求某点的函数值, 可以先在此做一记号, 再放上一个棋子, 用标有东, 南, 西, 北的正四面体骰子 (见图 5) 投掷, 如果落地的一面是东 (或南, 西, 北), 则向东 (或南, 西, 北) 走一步, 再掷再走, 一直到达边界为止. 这样便得到一条随机游动, 边界点的函数值即游动的随机变量的值 $\xi$. 进行充分多次的游动 (设为 $N$ 次), 记下 $\xi$ 对这 $N$ 次游动的值

$$\xi_1, \cdots, \xi_N.$$

其算术平均就是所欲求点的函数值的近似值.

**结论**　差分方法的误差由三部分构成: (i) 网格化时, 移动边界值所产生的误差. (ii) 差分化时, 把微商换成差分的误差. (iii) 解差分方程时, 代数法产生的是普通的误差, 而 Monte Carlo 法产生的是概率的误差.

因此, Monte Carlo 法的误差比代数法的误差更大些, 亦更不可靠些. 但另一方面, Monte Carlo 方法的计算程序特别简单, 而且如果我们只需要求得某些整点的函数值, 而不是全部整点的函数值, 用这一方法就更加经济了.

## 六　解析表达式 —— 有时会引入迷途

有些解析公式看来不错, 似乎是很解决问题的, 甚至于彻底解决问题的. 但如果不假思索地加以运用却会引入迷途. 如果较全面地理解"连续"与"离散"间的关系, 这些失误是完全可以避免的! 并且与此相反, 反而有相辅相成之妙, 也就是解析表达式可以启示新计算方法的苗头, 而不仅仅是理论上的重要性而已. 我们仍旧以 Laplace 方程的 Dirichlet 问题为例子, 并且取区域为单位圆. Laplace 方程的极坐标形式是

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial u}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2 u}{\partial\theta^2}=0. \tag{1}$$

**问题**　求连续函数 $u(\rho,\theta)$, 它在单位圆内适合 (1), 而在圆周 $\Gamma$ 上与已给的连续函数相符合. 即

$$u(\rho,\theta)|_{\rho=1}=\varphi(\theta). \tag{2}$$

今后常假定 $\varphi(\theta)$ 为 $[0,2\pi]$ 中有 $r(\geqslant 2)$ 阶连续微商, 而且是以 $2\pi$ 为周期的函数, 并且假定 $|\varphi^{(r)}(\theta)|<C$. 将 $\varphi(\theta)$ 展开成 Fourier 级数

$$\varphi(\theta)=\frac{a_0}{2}+\sum_{m=1}^{\infty}(a_m\cos m\theta+b_m\sin m\theta), \tag{3}$$

此处

$$a_m=\frac{1}{\pi}\int_0^{2\pi}\varphi(\theta)\cos m\theta\, d\theta,$$
$$b_m=\frac{1}{\pi}\int_0^{2\pi}\varphi(\theta)\sin m\theta\, d\theta. \tag{4}$$

容易看出

$$\rho^m\cos m\theta,\ \rho^m\sin m\theta(m=0,\ 1,\ 2,\ \cdots) \tag{5}$$

都是 (1) 的解, 而且分别以 $\cos m\theta$ 与 $\sin m\theta$ 为边界值. 因此可以希望

$$u(\rho,\theta)=\frac{a_0}{2}+\sum_{m=1}^{\infty}(a_m\cos m\theta+b_m\sin m\theta)\rho^m \tag{6}$$

为 (1) 适合 (2) 及 (3) 的解. 由于

$$a_m=O\left(\frac{1}{m^r}\right),\ b_m=O\left(\frac{1}{m^r}\right),$$

所以易见 (6) 的解是 (1) 适合 (2) 及 (3) 的解. 因为

$$\begin{aligned}\frac{1}{2}+\sum_{m=1}^{\infty}\rho^m\cos m\theta &= R\left(\frac{1}{1-\rho e^{i\theta}}\right)-\frac{1}{2}\\&=\frac{1-\rho\cos\theta}{1-2\rho\cos\theta+\rho^2}-\frac{1}{2}=\frac{1-\rho^2}{2(1-2\rho\cos\theta+\rho^2)},\end{aligned}$$

所以由 (4), (6) 得

$$\begin{aligned}&u(\rho,\theta)\\&=\frac{1}{\pi}\int_0^{2\pi}\varphi(\psi)\left[\frac{1}{2}+\sum_{m=1}^{\infty}(\cos m\theta\cos m\psi+\sin m\theta\sin m\psi)\rho^m\right]d\psi\\&=\frac{1}{\pi}\int_0^{2\pi}\varphi(\psi)\left[\frac{1}{2}+\sum_{m=1}^{\infty}\rho^m\cos m(\theta-\psi)\right]d\psi\\&=\frac{1}{2\pi}\int_0^{2\pi}\varphi(\psi)\frac{1-\rho^2}{1-2\rho\cos(\theta-\psi)+\rho^2}d\psi.\end{aligned}\tag{7}$$

这称为 Poisson 公式.

这是解答 $u(\rho,\theta)$ 的解析公式. 这公式的确很不错. 似乎都把问题彻底解决了. 但是仔细想一下, 是否真的解决问题了呢? 如果 $\varphi(\psi)$ 给了之后, 能够算出积分 (7)(即找到原函数), 则问题的确圆满解决了. 但如果算不出积分 (7)(这种情形比能算出的情形多得多), 或者当边界值仅仅由实验给出了若干数据时, 就产生了如何近似求解 $u(\rho,\theta)$ 的问题了. 很自然地会想到用数值积分的方法来近似计算 (7). 我们将在下面指出这样做会导出很荒谬的结论来.

(i) 矩形公式建议我们用

$$T_n(\rho,\theta)=\frac{1}{n}\sum_{l=0}^{n-1}\varphi\left(\frac{2\pi l}{n}\right)\frac{1-\rho^2}{1-2\rho\cos\left(\theta-\frac{2\pi l}{n}\right)+\rho^2}\tag{8}$$

来逼近 $u(\rho,\theta)$, 现在来看看当 $\rho\to1-0$ 时的情况:

$$\lim_{\rho\to1-0}T_n(\rho,\theta)=\begin{cases}0,\text{当}\theta\neq\dfrac{2\pi l}{n},\\\text{或}\theta=\dfrac{2\pi l}{n}\text{而}\varphi\left(\dfrac{2\pi l}{n}\right)=0\quad(0\leqslant l<n),\\\infty,\text{当}\theta=\dfrac{2\pi l}{n},\\\varphi\left(\dfrac{2\pi l}{n}\right)\neq0\quad(0\leqslant l<n).\end{cases}\tag{9}$$

因此用 $T_n(\rho,\theta)$ 来逼近 $u(\rho,\theta)$ 是十分荒谬的.

(ii) 我们在 Poisson 积分中, 用阶梯函数

$$\varphi^*(\theta)=\varphi\left(\frac{2\pi l}{n}\right),$$

其中

$$\frac{2\pi l}{n}\leqslant\theta<\frac{2\pi(l+1)}{n}\quad(0\leqslant l<n)\tag{10}$$

来代替 $\varphi(\theta)$. 换言之, 用

$$R_n(\rho,\theta)=\frac{1}{2\pi}\int_0^{2\pi}\frac{\varphi^*(\psi)(1-\rho^2)}{1-2\rho\cos(\psi-\theta)+\rho^2}d\psi\tag{11}$$

来逼近 $u(\rho,\theta)$. 由于

$$\log(1-\rho e^{i\theta})=-\sum_{m=1}^{\infty}\frac{(\rho e^{i\theta})^m}{m},$$

取虚部即得

$$\sum_{m=1}^{\infty}\rho^m\frac{\sin m\theta}{m}=\mathrm{tg}^{-1}\frac{\rho\sin\theta}{1-\rho\cos\theta}.$$

因此

$$\begin{aligned}R_n(\rho,\theta)=&\sum_{l=0}^{n-1}\frac{\varphi\left(\frac{2\pi l}{n}\right)}{2\pi}\int_{\frac{2\pi l}{n}}^{\frac{2\pi(l+1)}{n}}\frac{1-\rho^2}{1-2\rho\cos(\theta-\psi)+\rho^2}d\psi\\=&\frac{1}{n}\sum_{l=0}^{n-1}\varphi\left(\frac{2\pi l}{n}\right)\\&+\frac{1}{\pi}\sum_{l=0}^{n-1}\varphi\left(\frac{2\pi l}{n}\right)\sum_{m=1}^{\infty}\left[\sin m\left(\frac{2\pi(l+1)}{n}-\theta\right)\right.\\&\left.-\sin m\left(\frac{2\pi l}{n}-\theta\right)\right]\frac{\rho^m}{m}\\=&\frac{1}{n}\sum_{l=0}^{n-1}\varphi\left(\frac{2\pi l}{n}\right)+\frac{1}{\pi}\sum_{l=0}^{n-1}\left(\varphi\left(\frac{2\pi(l+1)}{n}\right)\right.\\&\left.-\varphi\left(\frac{2\pi l}{n}\right)\right)\mathrm{tg}^{-1}\frac{\rho\sin\left(\frac{2\pi l}{n}-\theta\right)}{1-\rho\cos\left(\frac{2\pi l}{n}-\theta\right)}.\end{aligned}\tag{12}$$

取 $\theta=C(1-\rho)^{\alpha}\left(\alpha>\frac{1}{2}\right)$. 则

$$\lim_{\rho\to1-0}\mathrm{tg}^{-1}\frac{\rho\sin\theta}{1-\rho\cos\theta}=\lim_{\rho\to1-0}\mathrm{tg}^{-1}C(1-\rho)^{\alpha-1}.$$

换言之, 当 $\rho \to 1-0,\ \theta \to 0$ 时, $\mathrm{tg}^{-1}\dfrac{\rho\sin\theta}{1-\rho\cos\theta}$ 可以趋近于 $\left[-\dfrac{\pi}{2},\dfrac{\pi}{2}\right]$ 中的任意值. 因此若 $\theta_0=\dfrac{2\pi l}{n}$, 而且 $\varphi\left(\dfrac{2\pi(l-1)}{n}\right)\neq\varphi\left(\dfrac{2\pi l}{n}\right)(0\leqslant l<n)$, 则当 $\rho\to 1-0,\ \theta\to\theta_0$ 时, $R_n(\rho,\theta)$ 的极限是不存在的. 所以必须给趋限的方法以限制. 例如规定趋限是延着向径的方向等. 而且可以证明, 虽然如此, 用 $R_n(\rho,\theta)$ 来逼近 $u(\rho,\theta)$, 精密度仍然是不高的. 在此就不作详细讨论了.

以上这两个从解析公式出发的近似计算方法都没有下面这个初等方法更为精密些.

设给了 $n(=2n'+1)$ 个点的函数值

$$y_l=\varphi\left(\frac{2\pi l}{n}\right)(|l|\leqslant n').$$

则如 (四) 所示. 命

$$S_n(\theta)=\sum_{m=-n'}^{n'}C'_m e^{im\theta},\qquad C'_m=\frac{1}{n}\sum_{l=-n'}^{n'}y_l e^{-2\pi ilm/n}\text{①}\tag{13}$$

如果 $\varphi(\theta)$ 在 $[-\pi,\pi]$ 中有 $r(\geqslant 2)$ 阶连续微商, 而且是周期为 $2\pi$ 的函数, 并且有 $|\varphi^{(r)}(\theta)|<C$, 则

$$|\varphi(\theta)-S_n(\theta)|<\frac{4C}{(r-1)n'^{r-1}}.\tag{14}$$

命

$$S_n(\rho,\theta)=\sum_{m=-n'}^{n'}C'_m e^{im\theta}\rho^{|m|}.\tag{15}$$

则

$$u(\rho,\theta)-S_n(\rho,\theta)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{(\varphi(\psi)-S_n(\psi))(1-\rho^2)}{1-2\rho\cos(\theta-\psi)+\rho^2}d\psi.$$

所以

$$|u(\rho,\theta)-S_n(\rho,\theta)|\leqslant\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|\varphi(\psi)-S_n(\psi)|(1-\rho^2)}{1-2\rho\cos(\theta-\psi)+\rho^2}d\psi$$

---

①为简单起见, 我们用复形式的 Fourier 级数. 复形式与实形式的 Fourier 级数的关系为

$$C_m=\frac{1}{2}(a_m-ib_m),$$
$$C_{-m}=\frac{1}{2}(a_m+ib_m)(m=1,2,\cdots).$$

$$
\begin{aligned}
&< \frac{4C}{(r-1)n'^{r-1}} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1-\rho^2}{1-2\rho\cos(\theta-\psi)+\rho^2} d\psi \\
&= \frac{4C}{(r-1)n'^{r-1}}. \qquad (16)
\end{aligned}
$$

在实际计算时, 因为

$$
\begin{aligned}
\sum_{m=-l}^{l} e^{imx}\rho^{|m|} &= \sum_{m=0}^{l}(e^{ix}\rho)^m + \sum_{m=0}^{l}(e^{-ix}\rho)^m - 1 \\
&= \frac{1-e^{i(l+1)x}\rho^{l+1}}{1-\rho e^{ix}} + \frac{1-e^{-i(l+1)x}\rho^{l+1}}{1-\rho e^{-ix}} - 1 \\
&= \frac{2-2\rho^{l+1}\cos(l+1)x - 2\rho\cos x + 2\rho^{l+2}\cos lx}{1-2\rho\cos x+\rho^2} - 1 \\
&= \frac{1-\rho^2-2\rho^{l+1}\cos(l+1)x+2\rho^{l+2}\cos lx}{1-2\rho\cos x+\rho^2},
\end{aligned}
$$

所以

$$
\begin{aligned}
S_n(\rho,\theta) &= \sum_{m=-n'}^{n'} \frac{1}{n} \sum_{l=-n'}^{n'} y_l e^{-2\pi ilm/n} e^{im\theta}\rho^{|m|} \\
&= \frac{1}{n}\sum_{l=-n'}^{n'} y_l \sum_{m=-n'}^{n'} e^{i\left(\theta-\frac{2\pi l}{n}\right)m}\rho^{|m|} \\
&= \frac{1}{n}\sum_{l=-n'}^{n'} y_l \frac{\left\{\begin{array}{c} 1-\rho^2-2\rho^{n'+1}\cos(n'+1)\left(\theta-\dfrac{2\pi l}{n}\right) \\ +2\rho^{n'+2}\cos n'\left(\theta-\dfrac{2\pi l}{n}\right)\end{array}\right\}}{1-2\rho\cos\left(\theta-\dfrac{2\pi l}{n}\right)+\rho^2}. \qquad (17)
\end{aligned}
$$

总之, 我们得到

**定理 1** 命 $u(\rho,\theta)$ 为方程 (1) 满足 (2) 的解, 此处 $\varphi(\theta)$ 为有 $r(\geqslant 2)$ 阶连续微商, 而且是有周期 $2\pi$ 的函数, 并且假定 $|\varphi^{(r)}(\theta)| < C$. 则

$$
\Bigg| u(\rho,\theta) - \frac{1}{n} \times \sum_{l=-n'}^{n'} \varphi\left(\frac{2\pi l}{n}\right)\left[\frac{1-\rho^2-2\rho^{n'+1}\cos(n'+1)\left(\theta-\dfrac{2\pi l}{n}\right)}{1-2\rho\cos\left(\theta-\dfrac{2\pi l}{n}\right)+\rho^2}\right.
$$

$$+\frac{2\rho^{n'+2}\cos n'\left(\theta-\frac{2\pi l}{n}\right)}{1-2\rho\cos\left(\theta-\frac{2\pi l}{n}\right)+\rho^{2}}\Bigg]\Bigg|<\frac{4C}{(r-1)n'^{r-1}}. \tag{18}$$

## 七 一致分布 —— 数论方法与 Monte Carlo 方法

要计算函数 $f(x)$ 在 [0,1] 上的积分, 我们可以把 [0,1] 分成 $n$ 等分, 取分点的函数值的算术平均, 用来作为 $f(x)$ 的积分的近似值 (矩形公式), 这就是化连续为离散的方法. 实际上, 不仅等分点有这样的性质, 凡是适合所谓"一致分布"条件的点列都有这个性质. 粗略地说, 一致分布的意义是说点列落在 [0, 1] 中任何一点附近的可能性都是相等的. 严格地, 可以定义如下:

命 $x_i(i=1,2,\cdots)$ 是 [0, 1] 间的一个点列, $a$ 为适合 $0\leqslant a\leqslant 1$ 的任意实数, $n$ 个点 $x_1,\cdots,x_n$ 落在分区间 [0,a] 中的个数用 $N_n(a)$ 表它. 如果常有

$$\lim_{n\to\infty}\frac{N_n(a)}{n}=a, \tag{1}$$

则称点列 $x_i(i=1,2,\cdots)$ 在 [0,1] 中一致分布.

关于一致分布有如下的判别条件.

**定理 1** 点列

$$x_1,\ \cdots,\ x_m,\ \cdots,\ 0\leqslant x_m\leqslant 1 \tag{2}$$

是一致分布的必要且充分的条件是对任一在 [0,1] 上可 Riemann 求积的函数 $f(x)$ 常有

$$\lim_{n\to\infty}\frac{f(x_1)+\cdots+f(x_n)}{n}=\int_0^1 f(x)dx. \tag{3}$$

**证** 先证明, 如果 $\{x_i\}$ 是一致分布, 则 (3) 式成立.

1) 取 $f(x)$ 是如下的函数

$$f(x)=\begin{cases}C, & 若 0\leqslant x\leqslant a,\\ 0, & 不然.\end{cases}$$

如此则

$$\lim_{n\to\infty}\frac{f(x_1)+\cdots+f(x_n)}{n}=C\lim_{n\to\infty}\frac{N_n(a)}{n}=Ca=\int_0^1 f(x)dx.$$

所以, 对于这样的函数 $f(x)$, 定理真实.

2) 如果 (3) 式对于 $f_1, \cdots, f_s$ 成立, 则对 $c_1 f_1 + \cdots + c_s f_s$ 也成立, 因此 (3) 式对所有的阶梯函数也真实.

3) 习知, 如果 $f$ 是一 Riemann 可积函数, 则任给 $\varepsilon > 0$, 皆有二阶梯函数 $\varphi_\varepsilon(x)$ 及 $\Phi_\varepsilon(x)$ 使

$$\varphi_\varepsilon(x) \leqslant f(x) \leqslant \Phi_\varepsilon(x),\ 0 \leqslant x \leqslant 1, \tag{4}$$

且使

$$\int_0^1 (\Phi_\varepsilon(x) - \varphi_\varepsilon(x))dx < \varepsilon. \tag{5}$$

由 2) 已知本定理对 $\Phi_\varepsilon(x)$ 及 $\varphi_\varepsilon(x)$ 真实, 所以

$$\begin{aligned}\int_0^1 \varphi_\varepsilon(x)dx &= \lim_{n\to\infty} \frac{1}{n}[\varphi_\varepsilon(x_1) + \cdots + \varphi_\varepsilon(x_n)] \\ &\leqslant \lim_{n\to\infty} \frac{1}{n}[f(x_1) + \cdots + f(x_n)] \\ &\leqslant \lim_{n\to\infty} \frac{1}{n}[\Phi_\varepsilon(x_1) + \cdots + \Phi_\varepsilon(x_n)] \\ &= \int_0^1 \Phi_\varepsilon(x)dx.\end{aligned}$$

故得

$$\left| \lim_{n\to\infty} \frac{f(x_1) + \cdots + f(x_n)}{n} - \int_0^1 f(x)dx \right| < \varepsilon.$$

这证明了定理的必要部分.

定理的充分部分的证明极为容易, 仅取

$$f(x) = \begin{cases} 1, & \text{若} 0 \leqslant x \leqslant a, \\ 0, & \text{不然}. \end{cases}$$

(3) 式就变为

$$\lim_{n\to\infty} \frac{N_n(a)}{n} = a.$$

定理证完.

显然, 一致分布的定义与它的判别条件可以很容易地推广至多个变数 (高维单位立方体) 的情况. 由定理 1 可见, 数值积分方法实依赖于一致分布点列的选取. 怎样选取最好的一致分布点列就是数值积分的中心问题. 习知, 对于计算 [0,1] 中的积分, 用等分点是能够导出最精密的误差的 (指误差的阶). 但在多变数的情况, 如果用等分点来进行计算, 误差依赖于积分的重数, 详细言之, 固定分点的个数, 则当积分的重数增加时, 误差亦随之而迅速增加. 或者可以说, 当要求有一定的精密度时, 则必需分点的数目随着积分重数的增加而迅速增加. 因此用这一方法来处理高

维空间的数值积分, 计算量十分巨大, 而难于实现. 具体地说, 对于 $s$ 重积分, 欲误差的精密度达到 $O(1/n)$, 则分点的个数需要 $O(n^s)$.

近年来发展起来的 Monte Carlo 方法, 是常用的高维空间的数值积分方法. 即随机地取 $n$ 个点 $(x_1^{(k)},\cdots,x_s^{(k)})$ $(k=1,2,\cdots,n)$, 然后以这 $n$ 个点的函数值的算术平均来逼近积分, 所谓"随机"的意思是指取每一点的概率都是相等的. 这样, 当 $n$ 充分大时, 就可能达到一定的精密度. 随机取点的方法一般都是在计算机上用数学方法来实现的. 而这些数学方法多为数论方法, 特别是同余式的方法. Monte Carlo 方法的优点在于在机器上运算的手续简便, 收敛速度虽然比矩形公式快些, 但是由这一方法得到的只能是概率的误差而不是真正的误差.

所谓数论方法, 即按照事先选定的最佳分布的点列上的函数值所构成的单和来逼近多重积分. 因而得到的误差不再是概率的, 而是肯定的, 不仅如此, 这些肯定的误差竟比概率误差还要好些, 而且可以证明, 对于某些函数类来说, 这种逼近的误差的主阶已经臻于至善了. 具体地说, 误差的主阶与单重积分是一样的.

**例** 　假定 $f(x_1,\cdots,x_5)$ 为各变数皆有二阶连续微商的函数. 且各阶微商皆为各变数有周期 1 的函数, 且

$$\left|\frac{\partial^r f}{\partial x_1^{i_1}\cdots\partial x_5^{i_5}}\right| < C(2\pi)^r(i_1+\cdots+i_5=r,\ 0\leqslant r\leqslant 10,\ 2\geqslant i_j\geqslant 0).$$

则

$$\begin{aligned}&\Bigg|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_5)dx_1\cdots dx_5\\&\quad-\frac{1}{15019}\sum_{k=1}^{15019} f\left(\frac{k}{15019},\frac{10641k}{15019},\frac{2640k}{15019},\frac{6710k}{15019},\frac{781k}{15019}\right)\Bigg|\\&<0.0032\left(\frac{\pi^2}{6}\right)^5 C.\end{aligned}$$

必需指出, 数论方法不仅在数值积分方面有用, 而且可以用于函数逼近论及积分方程的渐近求解等方面. 例如, 我们可以证明, 适合某些光滑条件的各变数皆有周期为 1 的函数, 都可以用一个三角多项式来逼近, 而逼近的主阶不依赖于函数的变数的个数. 关于这些方面, 请参看 [1].

[1] 华罗庚, 王元, 数值积分及其应用, 科学出版社, 1963.

[2] 华罗庚, 高等数学引论, 科学出版社, 1963.

(原载 1963 年 1 月"科学通报")

# 关于在等高线图上计算矿藏储量与坡地面积的问题①

## §1 引 言

感谢我国的地理、矿冶与地质工作者们, 他们向我们介绍了不少计算矿藏储量与计算坡地面积的实用方法, 使我们能学习到这些方法, 从而进行了一些研究. 作者试图在本文中对这些方法进行比较, 阐明它们相互之间的关系, 与这些方法的偏差情况, 并提出若干建议.

关于分层计算矿藏储量方面, 在矿体几何学上 (见 [2] ~ [4]) 有Бауман 公式, 截锥公式与梯形公式. 设用它们算出来的矿藏体积分别为 $v$, $v_1$ 与 $v_2$. 本文证明了它们满足不等式:

$$v \leqslant v_1 \leqslant v_2,$$

并且完全确定了取等号的情况. 关于这三个公式的比较问题, 作者认为主要应从量纲来看, 因此我们认为Бауман 公式的局限性较少.

本文提供了一个双层合算矿藏储量的公式, 这个公式的获得首先在于我们找到了Бауман 公式的一个新证明. 这个证明既简单, 而又易于进一步改进. 它的优点在于比Бауман 公式麻烦得并不很多, 但比Бауман 公式多考虑了一些因素, 同时也比Соболевский 公式 (即通常的双层合算矿藏储量的公式, 见 [2] ~ [4]) 多考虑了一些因素. 我们推荐它供我国矿藏储量计算工作者参考或试用.

关于坡地面积的计算方面, 在地理学上常用 Волков 方法 (见 [5]~[6]); 在矿体几何学上, 则常用Бауман方法 (见 [1]~[2]). 本文指出, Бауман 方法比 Волков 方法精密, 但用这两个方法算出的结果常比真正的结果偏低. 本文完全定出了能够用这两个方法来无限精密地计算其面积的曲面及指出这两个方法的偏差情况. 详言之, 偏差依赖于曲面上点的倾角的变化. 只有当整个曲面上各点的倾角都相差不大时, Волков 方法才能得到精确结果, 而只有当曲面在相邻两等高线间的点的倾角的变化不大时, Бауман 方法才能给出精密的结果, 然而在其他情况下, 用这两个方法的误差就可能比较大了, 因此我们建议在等高线图上通过制高点引进若干条放射线, 当曲面与直纹面相近时, 可以分别求出相邻两条放射线间的表面积, 然后总加起来. 如果相邻两条等高线间与相邻两条放射线间, 曲面的倾角的变化都比较大时,

① 与王元同志合作.

可以分别算出由放射线及等高线所织成的每一小块的表面积, 然后总加起来. 这样算出的结果, 偏差就比较小了.

## §2　矿藏储量计算

**1. Бауман 方法**

假定有一张矿藏的等高线图, 高程差是 $h$, 地图上所表示的一圈, 实际上便是一定高程的矿体的截面积. 我们来估计两张这样的平面之间的矿藏的体积. 这两张平面之间的距离便是高程差 $h$. 我们以 $A, B$ 各表示下、上两个等高线圈所包围的截面 (见图 1, 它们的面积亦记为 $A, B$). Бауман建议用

$$v = \left[\frac{1}{2}(A+B) - \frac{T(A,B)}{6}\right]h \tag{1}$$

来估算这两个高程间的一片的体积 $v$, 此处 $T(A,B)$ 是用以下方法所画出的图形的面积, 称它为Бауман 改正数.

如图 2 中, 从制高点 $O$ 出发, 作放射线 $OP$, 这放射线在地图上 $A, B$ 之间的长度是 $l$. 另作图 3, 取一点 $O'$, 与 $OP$ 同方向取 $O'P' = l$. 当 $P$ 延着 $A$ 的周界走一圈时, $P'$ 也得一图形, 这图形的面积就称为Бауман改正数. 因为它依赖于两截面 $A$ 与 $B$, 所以我们用 $T(A,B)$ 来表示它.

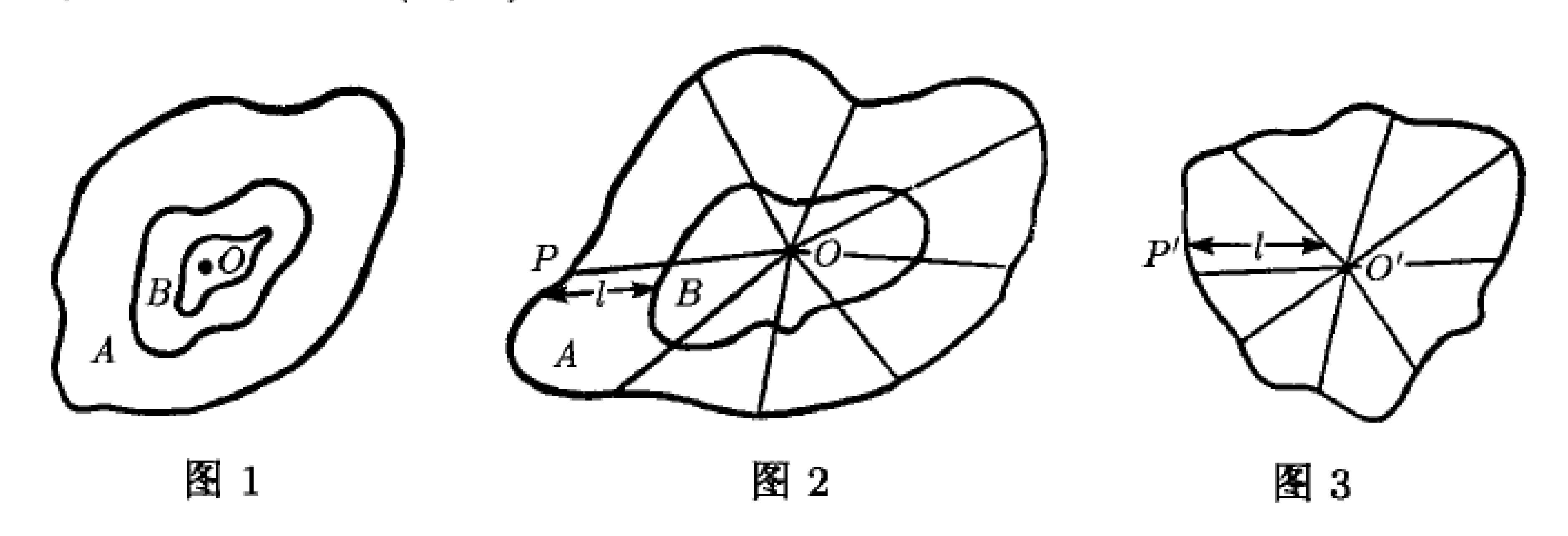


图 1　　图 2　　图 3

把算出来的矿体体积一片一片地加起来, 就得到矿藏的体积 $V$. 换言之, 设矿体的等高线图的 $n+1$ 条等高线所围成的面积依次为 $S_0, S_1, \cdots, S_n$, 则矿体的体积 $V$ 由下式来近似计算:

$$V = \left(\frac{S_0+S_n}{2} + \sum_{m=1}^{n-1} S_m\right)h - \frac{h}{6}\sum_{m=0}^{n-1} T(S_m, S_{m+1}), \tag{2}$$

此处 $h$ 为高程差 (图 4).

**定理**　(Бауман) 已知物体的下底 $A$ 与上底 $B$(其面积亦记为 $A, B$) 均为平面, 且 $A$ 平行于 $B$,$h$ 为它们之间的高, $O$ 为 $B$ 上一点, 若用任意通过 $O$ 而垂直于 $B$ 的平面来截物体, 所得的截面都是四边形, 则物体的体积 $v$ 恰如 (1) 式所示.

**证** 以 $O$ 为中心, 引进极坐标 (见图 5).

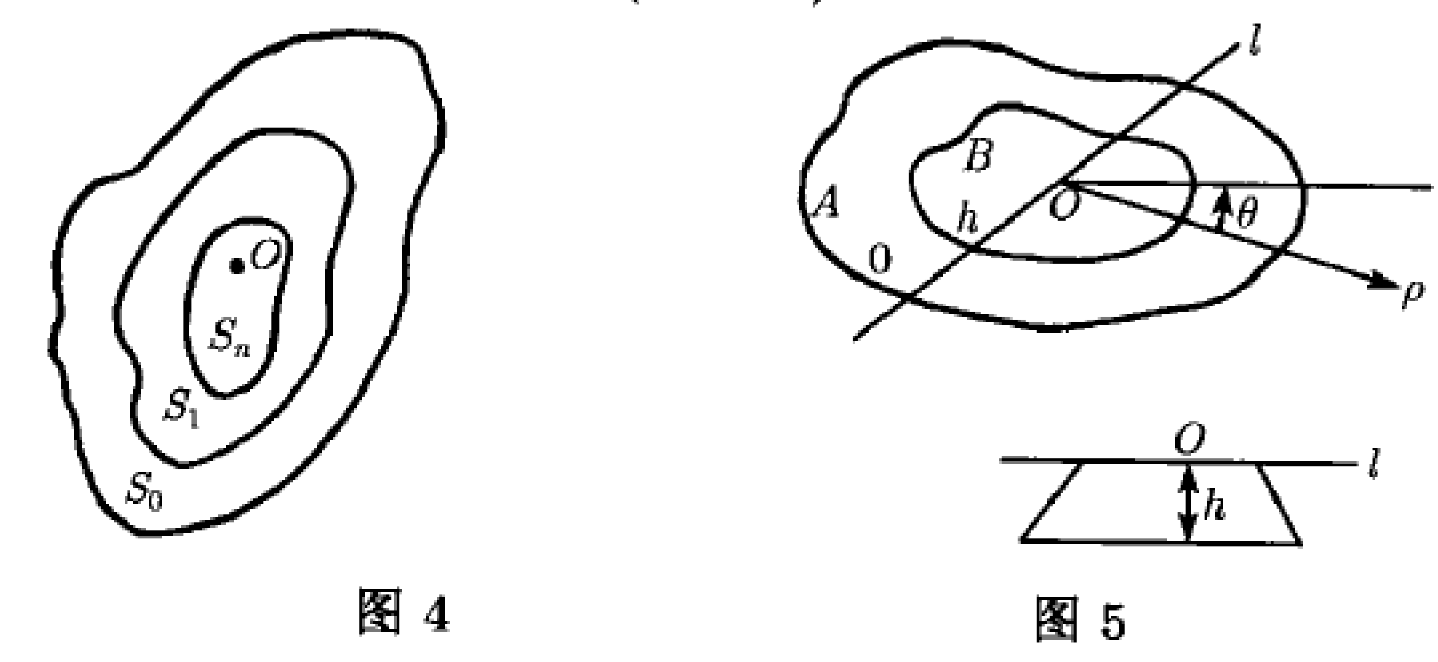



图 4 图 5

命高度为 $z$ 的等高线的极坐标方程为

$$\rho = \rho(z, \theta) \quad (0 \leqslant \theta \leqslant 2\pi),$$

其中, $\rho(z, 0) = \rho(z, 2\pi)$. 今后我们常假定 $\rho(z, \theta)$ $(0 \leqslant \theta \leqslant 2\pi, 0 \leqslant z \leqslant h)$ 是连续的, 我们不妨假定 $A, B$ 的高程各为 0 及 $h$. 并且记

$$\rho_1(\theta) = \rho(0, \theta), \quad \rho_2(\theta) = \rho(h, \theta).$$

由假定可知

$$\rho(z, \theta) = \frac{z}{h}\rho_2(\theta) + \frac{h-z}{h}\rho_1(\theta) \quad (0 \leqslant z \leqslant h).$$

因此物体的体积为

$$\begin{aligned}
&\frac{1}{2}\int_0^h \int_0^{2\pi} \rho^2(z, \theta) d\theta \, dz \\
=&\frac{1}{2}\int_0^{2\pi} \int_0^h \left(\frac{z}{h}\rho_2(\theta) + \frac{h-z}{h}\rho_1(\theta)\right)^2 dz \, d\theta \\
=&\frac{h}{2}\int_0^{2\pi} \left(\frac{\rho_1^2(\theta)}{3} + \frac{\rho_2^2(\theta)}{3} + \frac{\rho_1(\theta)\rho_2(\theta)}{3}\right) d\theta \\
=&\frac{h}{2}\left[\frac{1}{2}\int_0^{2\pi} \rho_1^2(\theta) d\theta + \frac{1}{2}\int_0^{2\pi} \rho_2^2(\theta) d\theta\right] \\
&- \frac{h}{6}\left[\frac{1}{2}\int_0^{2\pi} (\rho_1(\theta) - \rho_2(\theta))^2 d\theta\right] \\
=&\frac{h}{2}(A+B) - \frac{h}{6}T(A, B).
\end{aligned}$$

定理证完.

**2. Бауман 公式, 截锥公式与梯形公式的关系**

假定物体的下底 $A$ 与上底 $B$ 均为平面, 且 $A$ 平行于 $B, h$ 为它们之间的高, $O$ 为 $B$ 上一点, 除Бауман 公式外, 常用下面两公式来近似计算物体的体积:

截锥公式:

$$v_1 = \frac{h}{3}(A+B+\sqrt{AB}), \tag{3}$$

梯形公式:

$$v_2 = \frac{h}{2}(A+B), \tag{4}$$

通常当 $\dfrac{A-B}{A} > 40\%$ 时, 用公式 (3), 而当 $\dfrac{A-B}{A} < 40\%$ 时, 用公式 (4).

**定理 1**　不等式

$$v \leqslant v_1 \leqslant v_2 \tag{5}$$

恒成立, 当且仅当物体为截锥, 且此锥体的顶点至底面 $A$ 的垂线通过点 $O$ 时, $v = v_1$, 当且仅当 $A = B$ 时, $v_1 = v_2$.

**证**　如Бауман 定理中的假定. 由Бауман 公式及Буняковский–Schwarz 不等式可知

$$\begin{aligned} v =& \frac{h}{6}\int_0^{2\pi}(\rho_1^2(\theta)+\rho_2^2(\theta)+\rho_1(\theta)\rho_2(\theta))d\theta \\ \leqslant& \frac{h}{3}\left[\frac{1}{2}\int_0^{2\pi}\rho_1^2(\theta)d\theta + \frac{1}{2}\int_0^{2\pi}\rho_2^2(\theta)d\theta \right. \\ &\left. + \frac{1}{2}\sqrt{\int_0^{2\pi}\rho_1^2(\theta)d\theta\int_0^{2\pi}\rho_2^2(\theta)d\theta}\right] \\ =& \frac{h}{3}(A+B+\sqrt{AB}) = v_1, \end{aligned}$$

当且仅当 $\rho_1(\theta) = c\rho_2(\theta)$ $(0 \leqslant \theta \leqslant 2\pi$, $c$ 为常数) 时, 即当这物体为一截头锥体, 而此锥体的顶点至底面 $A$ 的垂线通过点 $O$ 时, 才会取等号 (图 6).

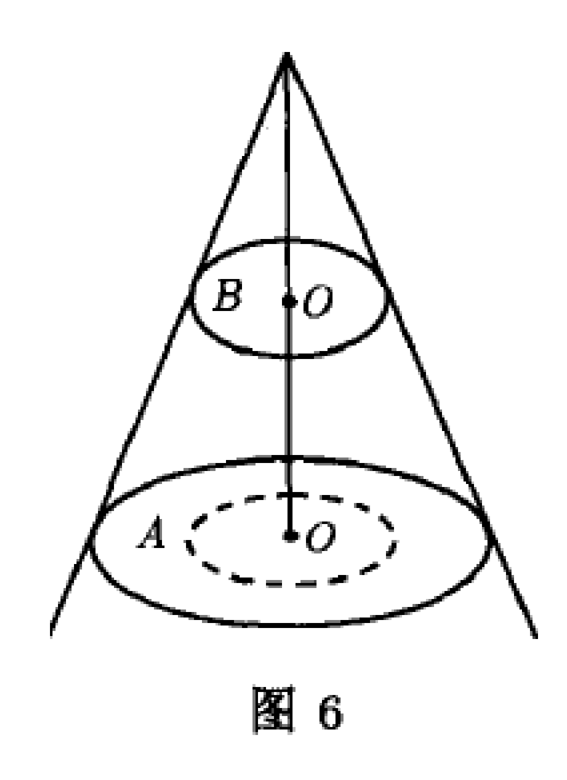



图 6

又由于

$$v_2 - v_1 = \frac{h}{2}(A+B) - \frac{h}{3}(A+B+\sqrt{AB})$$

$$=\frac{h}{6}(\sqrt{A}-\sqrt{B})^2 \geqslant 0,$$

所以

$$v_1 \leqslant v_2$$

当且仅当 $A=B$ 时取等值, 定理证完.

关于这三个公式的比较问题, 我们认为主要应该从量纲来看, 面的量纲为 2. 所以把面的量纲考虑为 1 所得出的公式, 局限性往往是比较大的.

梯形公式是把中间截面看成上底与下底的算术平均而得到的, 所以把面的量纲当作 1.

Бауман 公式则是将中间截面作为量纲 2 来考虑的. 详言之, 它假定了 $\rho(z,\theta)$ 为 $\rho(0,\theta)$ 与 $\rho(h,\theta)$ 关于 $z$ 的线性关系而得到的 (见 1).

截锥公式亦是将中间截面的量纲考虑为 2. 但比Бауман 公式还多假定了 $\rho(0,\theta)=c\rho(h,\theta)(0\leqslant\theta\leqslant 2\pi)$, 此处 $c$ 为一常数.

因此我们认为Бауман 公式更具有普遍性, 所以用它来近似计算物体的体积, 一般说来, 应该比较精确, 但这并不排斥对于某些个别物体, 用其他两个公式更恰当些的可能性. 例如有一梯形, 其上底与下底的宽度相等 (如图 7 所示). 用梯形公式反而能获得它的真正体积, 而用Бауман 公式与截锥公式来计算, 结果就偏低了. 不过, 我们注意此时这梯形的截面的量纲为 1(由于沿 $y$ 轴未变).

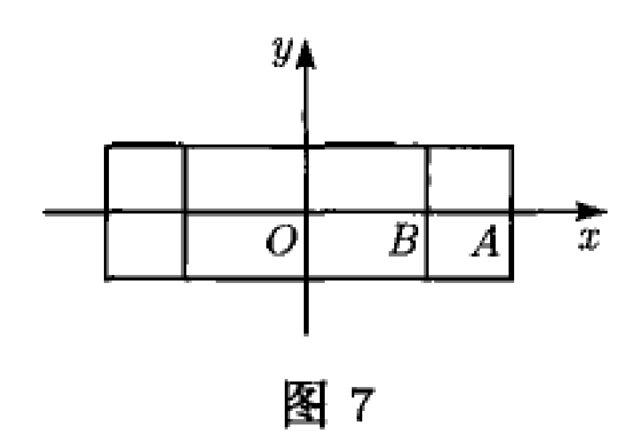



图 7

相对于Бауман 公式, 我们还可以估计用梯形公式与截锥公式的相对偏差.

例如当 $\dfrac{A-B}{A}<40\%$ $\left(即 B>\dfrac{3}{5}A\right)$ 时, 用梯形公式算出的结果相对于Бауман 公式算出的结果的相对偏差为

$$\varDelta=\frac{v_2-v}{v}=\frac{\frac{1}{2}(A+B)h-\frac{1}{2}(A+B)h+\frac{h}{6}T(A,B)}{\frac{1}{2}(A+B)h-\frac{h}{6}T(A,B)}$$

$$=\frac{T(A,B)}{3(A+B)-T(A,B)}.$$

因为 $T(A,B)\leqslant A-B$(即

$$\frac{1}{2}\int_0^{2\pi}(\rho_1(\theta)-\rho_2(\theta))^2d\theta\leqslant\frac{1}{2}\int_0^{2\pi}\rho_1^2(\theta)d\theta-\frac{1}{2}\int_0^{2\pi}\rho_2^2(\theta)d\theta,$$

此不等式显然成立), 所以

$$\Delta \leqslant \frac{A-B}{2A+4B}.$$

再以条件 $B > \frac{3}{5}A$ 代入, 得

$$\Delta \leqslant \frac{A-\frac{3}{5}A}{2A+\frac{12}{5}A} = \frac{1}{11} < 10\%.$$

**3. 建议一个计算矿藏储量的公式**

Бауман 公式是假定 $\rho(z,\theta)$ 为 $\rho(0,\theta)$ 与 $\rho(h,\theta)$ 关于 $z$ 的线性关系而得到的. 如果我们将两相邻分层放在一起估计, 即已知相邻三等高线 $\rho(0,\theta),\rho(h,\theta)$ 与 $\rho(2h,\theta)$. 我们用通过 $\rho(0,\theta),\rho(h,\theta)$ 与 $\rho(2h,\theta)$ 的抛物线所形成的曲面 $\rho=\rho(z,\theta)$ 来逼近矿体这两分层的表面, 因此我们建议用如下的计算方法.

命 $A,B,C$ 分别表示连续三等高线所围成的截面 (面积亦记为 $A,B,C$), $A$ 与 $B$ 及 $B$ 与 $C$ 之间的距离都是 $h$, 则这两片在一起的体积可用以下公式来近似计算

$$\begin{aligned} v_3 =& \frac{h}{3}(A+4B+C) - \frac{h}{15}(2T(A,B) \\ &+ 2T(B,C) - T(A,C)). \end{aligned} \tag{6}$$

如果不计 (6) 式中的第二项, 就是熟知的Соболевский 公式. 把二片二片的体积总加起来, 就得到矿藏的总体积 $V$ 的近似公式. 换言之, 设矿藏的等高线图的 $2n+1$ 条等高线所围成的面积依次为 $S_0,S_1,\cdots,S_{2n}$, 而高程差为 $h$, 则矿藏的体积 $V$ 由下式来近似计算

$$\begin{aligned} V =& \frac{h}{3}\left[S_0+S_{2n}+4\sum_{i=0}^{n-1}S_{2i+1}+2\sum_{i=0}^{n-1}S_{2i}\right] \\ &-\frac{h}{15}\left[2\sum_{i=0}^{n-1}T(S_{2i},S_{2i+1})+2\sum_{i=0}^{n-1}T(S_{2i+1},S_{2i+2})\right. \\ &\left.-\sum_{i=0}^{n-1}T(S_{2i},S_{2i+2})\right]. \end{aligned} \tag{7}$$

注意: 如果等高线图含有偶数条等高线, 则最上面一片可以单独估计, 其余的用公式 (7).

**定理 2** 已知物体的上底 $C$ 与下底 $A$ 均为平面, $B$ 为中间截面 (面积亦分别记为 $C,A,B$), 且 $A,C$ 都与 $B$ 平行, $A$ 与 $B$ 之间及 $B$ 与 $C$ 之间的距离都是 $h$, $O$ 为 $C$ 上一点 (图 8). 若用任意通过 $O$ 而垂直于 $C$ 的平面截物体, 所得的截面的周界均由两条直线及两条抛物线所构成, 则物体的体积 $v_3$ 恰如 (6) 式所示.

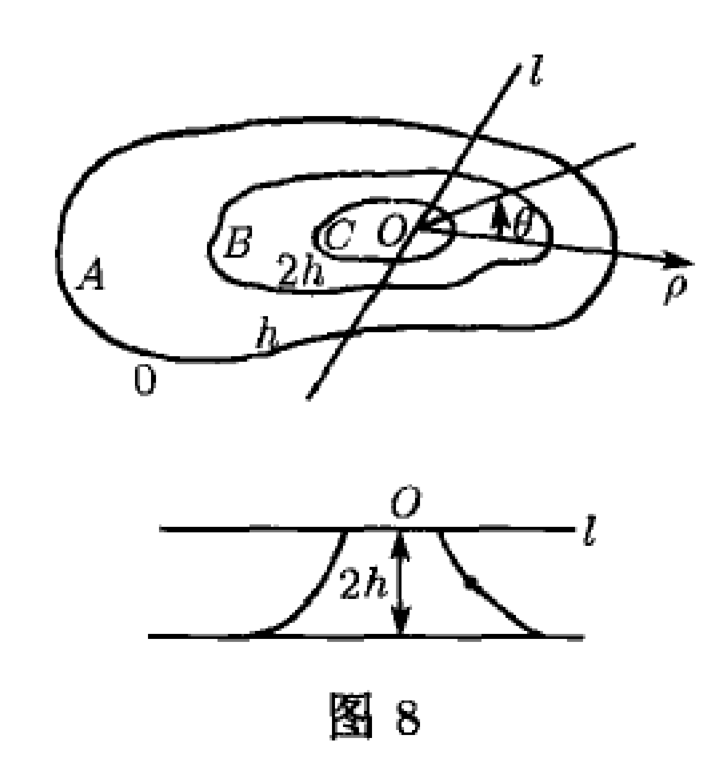



图 8

**证** 以 $O$ 为中心, 引进极坐标, 命高度为 $z$ 的等高线的极坐标方程为

$$\rho=\rho(z,\theta)(0\leqslant\theta\leqslant 2\pi,\ \rho(z,0)=\rho(z,2\pi)).$$

不妨假定 $A,B,C$ 的高程分别为 $0,h,2h$, 并且记

$$\rho_1(\theta)=\rho(0,\theta),\ \rho_2(\theta)=\rho(h,\theta),\ \rho_3(\theta)=\rho(2h,\theta)$$

由假定可知

$$\begin{aligned}\rho(z,\theta)=&\frac{(z-h)(z-2h)}{2h^2}\rho_1(\theta)\\&-\frac{z(z-2h)}{h^2}\rho_2(\theta)+\frac{z(z-h)}{2h^2}\rho_3(\theta)\end{aligned}\tag{8}$$

因此物体的体积 $v_3$ 为

$$\begin{aligned}&\frac{1}{2}\int_0^{2h}\int_0^{2\pi}\rho^2(z,\theta)d\theta\ dz\\=&\frac{1}{2}\int_0^{2\pi}d\theta\int_0^{2h}\left[\frac{(z-h)(z-2h)}{2h^2}\rho_1(\theta)\right.\\&\left.-\frac{z(z-2h)}{h^2}\rho_2(\theta)+\frac{z(z-h)}{2h^2}\rho_3(\theta)\right]^2dz\\=&\frac{h}{2}\int_0^{2\pi}\left[\frac{4}{15}\rho_1^2(\theta)+\frac{16}{15}\rho_2^2(\theta)+\frac{4}{15}\rho_3^2(\theta)+\frac{4}{15}\rho_1(\theta)\rho_2(\theta)\right.\\&\left.+\frac{4}{15}\rho_2(\theta)\rho_3(\theta)-\frac{2}{15}\rho_1(\theta)\rho_2(\theta)\right]d\theta\\=&\frac{h}{2}\int_0^{2\pi}\left[\frac{\rho_1^2(\theta)}{3}+\frac{4\rho_2^2(\theta)}{3}+\frac{\rho_3^2(\theta)}{3}\right.\\&-\frac{2}{15}(\rho_1(\theta)-\rho_2(\theta))^2\frac{2}{15}(\rho_2(\theta)-\rho_3(\theta))^2\\&\left.+\frac{1}{15}(\rho_1(\theta)-\rho_2(\theta))^2\right]d\theta\end{aligned}$$

$$=\frac{h}{3}(A+4B+C)-\frac{h}{15}(2T(A,B)+2T(B,C)-T(A,C)).$$

定理证完.

## §3 坡地面积计算

**4.** Бауман**方法及** Волков **方法**

现在先介绍矿学家及地理学家所常用的方法, 假定地图上以 $\Delta h$ 为高程差画出等高线, 今后我们常假定有一制高点, 及等高线成圈的情况来讨论 (其他情况也可以十分容易地被推出来). 我们假定由制高点出发, 向外一圈一圈地画出等高线 $(l_{n-1})$, $(l_{n-2})$, $\cdots$, $(l_0)$ (图 9). 记 $(l_0)$ 的高度为 0, 而制高点用 $(l_n)$ 表之, 它的高度是 $h$, $(l_i)$ 与 $(l_{i+1})$ 之间的面积用 $B_i$ 表示 (即投影的面积).

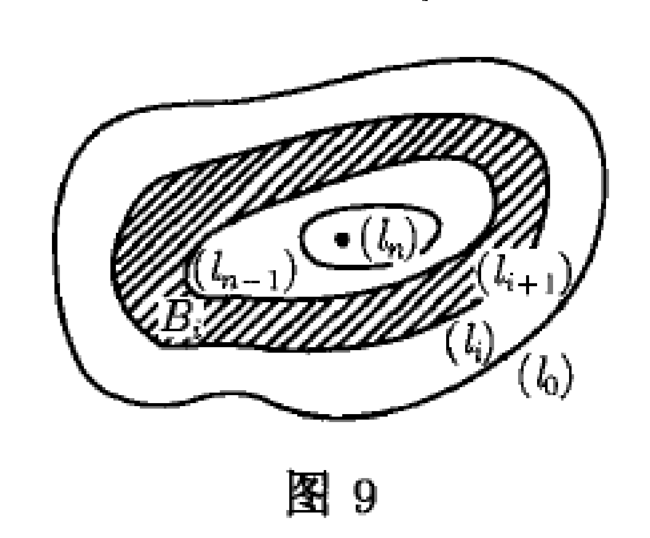



图 9

I. 矿体几何学上常用的方法的步骤如下:

a. $C_i=\frac{1}{2}(l_i+l_{i+1})\Delta h$ (中间直立隔板的面积);

b. $\sum_{i=0}^{n-1}\sqrt{B_i^2+C_i^2}$ 就是所求的斜面积的渐近值 (Бауман 方法).

II. 地理学上常用的方法的步骤如下:

a. $l=\sum_{i=0}^{n-1}l_i$ (等高线的总长度), $B=\sum_{j=0}B_j$ (总投影面积), $\operatorname{tg}\alpha=\frac{\Delta h\cdot l}{B}$(平均倾角);

b. $B\sec\alpha=\sqrt{B^2+(\Delta h\cdot l)^2}$ 就是所求的斜面积的渐近式 (Волков 方法).

**附记** $\sqrt{a^2+b^2}$ 可以借商高定理, 用图解法很快求出.

这两个方法哪一个更好一些? 这些方法给出的结果在怎样的程度上追近斜面积? 换句话说, 当等高线的分布趋向无限精密时 (也就是 $\Delta h\to 0$ 时), 这些方法所给出的结果是什么? 是否就是真正的斜面积呢? 一般说来, 答案是否定的, 仅仅是一些十分特殊的曲面, 答案才是肯定的. 我们将在下面定出这些曲面, 并将给出这些方法和实际结果的相差比例, 同时指出避免较大偏差的计算步骤.

**5.** Бa, **B0 与**$S$**的关系**

以制高点 $(l_n)$ 为中心 $O$, 引进极坐标. 命高度为 $z$ 的等高线方程为

$$\rho=\rho(z,\theta)\quad(0\leqslant\theta\leqslant 2\pi),$$

其中 $\rho(z,\theta)=\rho(z,2\pi)$. 我们在今后常假定 $\dfrac{\partial\rho(z,\theta)}{\partial\theta}$ 与 $\dfrac{\partial\rho(z,\theta)}{\partial z}$ $(0\leqslant\theta\leqslant 2\pi,\ 0\leqslant z\leqslant h)$ 都是连续的. 命 $z_i=\dfrac{h}{n}i$, 则 $l_i$ 所包围的面积等于

$$\frac{1}{2}\int_0^{2\pi}\rho^2(z_i,\theta)d\theta,$$

所以由中值公式可知

$$\begin{aligned}B_i=&\frac{1}{2}\int_0^{2\pi}[\rho^2(z_i,\theta)-\rho^2(z_{i+1},\theta)]d\theta\\=&-\int_0^{2\pi}\rho(z_i',\theta)\frac{\partial\rho(z_i',\theta)}{\partial z_i'}d\theta\ \Delta h,\end{aligned}$$

此处 $z_i'\in[z_i,z_{i+1}]$, 而 $\Delta h=\dfrac{h}{n}$. 另一方面, $(l_i)$ 的长度等于

$$l_i=\int_{(l_i)}ds=\int_0^{2\pi}\sqrt{\rho^2(z_i,\theta)+\left(\frac{\partial\rho(z_i,\theta)}{\partial\theta}\right)^2}d\theta.$$

由Bayман 方法所得出的结果是

$$C_i=\int_0^{2\pi}\sqrt{\rho^2(z_i'',\theta)+\left(\frac{\partial\rho(z_i'',\theta)}{\partial\theta}\right)^2}d\theta\ \Delta h,$$

这里用了中值公式, $z_i''\in[z_i,z_{i+1}]$, 因此当 $\Delta h\to 0$ 时,

$$\sum_{i=0}^{n-1}\sqrt{B_i^2+C_i^2}$$

趋近于

$$\mathrm{Ba}=\int_0^h\sqrt{\left(\int_0^{2\pi}\rho\frac{\partial\rho}{\partial z}d\theta\right)^2+\left(\int_0^{2\pi}\sqrt{\rho^2+\left(\frac{\partial\rho}{\partial\theta}\right)^2}d\theta\right)^2}dz.\qquad(9)$$

这便是用Bayман 方法算出的斜面积, 当 $\Delta h\to 0$ 时所趋向的数值.

又易见

$$B=\frac{1}{2}\int_0^{2\pi}\rho^2(0,\theta)d\theta=\int_0^{2\pi}d\theta\int_0^h-\rho\frac{\partial\rho}{\partial z}dz$$

(注意: $\rho(h,\theta)=0$) 及 $\Delta hl$ 的极限应当等于

$$\lim_{n\to\infty}\sum_{i=0}^{n-1}\frac{h}{n}\int_0^{2\pi}\sqrt{\rho^2(z_i,\theta)+\left(\frac{\partial\rho(z_i,\theta)}{\partial\theta}\right)^2}d\theta$$
$$=\int_0^h dz\int_0^{2\pi}\sqrt{\rho^2+\left(\frac{\partial\rho}{\partial\theta}\right)^2}d\theta,$$

因此用 Волков 方法算出的斜面积, 当 $\Delta h\to 0$ 时, 所趋向的数值为

$$\mathrm{B0}=\sqrt{\left(\int_0^{2\pi}d\theta\int_0^h-\rho\frac{\partial\rho}{\partial z}dz\right)^2+\left(\int_0^{2\pi}d\theta\int_0^h\sqrt{\rho^2+\left(\frac{\partial\rho}{\partial\theta}\right)^2}dz\right)^2}. \tag{10}$$

由于

$$ds^2=\left[\left(\frac{\partial\rho}{\partial\theta}\right)^2+\rho^2\right]d\theta^2+2\frac{\partial\rho}{\partial\theta}\frac{\partial\rho}{\partial z}d\theta\ dz+\left(1+\left(\frac{\partial\rho}{\partial z}\right)^2\right)dz^2,$$

所以斜面的面积 $S$ 为

$$S=\int_0^{2\pi}d\theta\int_0^h\sqrt{\rho^2+\left(\frac{\partial\rho}{\partial\theta}\right)^2+\left(-\rho\frac{\partial\rho}{\partial z}\right)^2}d\theta. \tag{11}$$

为了比较Бa, Bo 与 $S$, 我们引进一个复值函数

$$f(z,\theta)=\rho\frac{\partial\rho}{\partial z}+i\sqrt{\rho^2+\left(\frac{\partial\rho}{\partial\theta}\right)^2}, \tag{12}$$

则得

$$\mathrm{Ba}=\int_0^h\left|\int_0^{2\pi}f(z,\theta)d\theta\right|dz, \tag{13}$$

$$\mathrm{B0}=\left|\int_0^h\int_0^{2\pi}f(z,\theta)d\theta dz\right|, \tag{14}$$

及

$$S=\int_0^h\int_0^{2\pi}|f(z,\theta)|d\theta dz. \tag{15}$$

因此显然有不等式

$$\mathrm{B0}\leqslant\mathrm{Ba}\leqslant S. \tag{16}$$

由此可见: (i) Бayман 方法比 Волков 方法精密; (ii) 所求出的结果比真正的结果偏低一些; (iii) Бayман 方法既然偏低, 因此可以作如下的修改, 即取 $C_i=l_i\Delta h$. 这样既简化了算法而又增大了数值.

现在来考虑 $\mathrm{B0}=S$ 及Бa$=S$ 的曲面, 先讲下面的引理:

**引理** 若 $f(x)$ 为区间 $[a,b]$ 中的复值函数, 此处 $a,b$ 均为实数, 则等式

$$\left|\int_a^b f(x)dx\right| = \int_a^b |f(x)|dx \tag{17}$$

成立的必要且充分的条件是 $f(x)$ 的虚实部分之比为常数.

**证** 命 $f(x)=\rho(x)e^{i\theta(x)}, \rho(x)\geqslant 0$, 而 $\theta(x)$ 是实函数. 显然如果 $\theta(x)$ 为与 $x$ 无关的常数, 则 (17) 成立. 反之, 由于

$$\begin{aligned}\left(\left|\int_a^b f(x)dx\right|\right)^2 &= \int_a^b\int_a^b f(x)\overline{f(y)}dx\ dy\\ &= \int_a^b\int_a^b \rho(x)\rho(y)e^{i(\theta(x)-\theta(y))}dx\ dy\\ &=2\iint_{a\leqslant x<y\leqslant b}\rho(x)\rho(y)\cos[\theta(x)-\theta(y)]dx\ dy,\end{aligned}$$

$$\left(\int_a^b |f(x)|dx\right)^2 = 2\iint_{a\leqslant x<y\leqslant b}\rho(x)\rho(y)dx\ dy$$

因而若 (17) 成立, 则必

$$\cos(\theta(x)-\theta(y))\equiv 1,$$

即 $\theta(x)\equiv\theta(y)$. 此即引理所需.

易知对于多重积分, 引理依然成立.

由引理可知

$$\text{B0} = \left|\int_0^{2\pi}\int_0^h f(z,\theta)dx\ d\theta\right| = \int_0^{2\pi}\int_0^h |f(z,\theta)|dx\ d\theta = S$$

成立的必要且充分的条件为 $f(z,\theta)$ 的虚实部分之比是常数 $c$, 则得偏微分方程

$$\rho^2+\left(\frac{\partial\rho}{\partial\theta}\right)^2 = c^2\left(\rho\frac{\partial\rho}{\partial z}\right)^2. \tag{18}$$

换言之, 仅有适合这偏微分方程的函数 $\rho=\rho(z,\theta)$, Волков 方法才能给出正确答案. 这当然要适合以下的条件: $\rho(h,\theta)=0$ (这是制高点) 及 $\rho(0,\theta)=\rho_0(\theta)$ (这是曲面的底盘方程).

我们并不解这偏微分方程, 而从它的几何意义入手, 把 $\theta$ 与 $z$ 看成参变数, 即

$$x=\rho\cos\theta,\quad y=\rho\sin\theta,\quad z=z,$$

而 $\rho$ 是 $\theta$ 与 $z$ 的函数, 由

$$\frac{\partial x}{\partial\theta}=\frac{\partial\rho}{\partial\theta}\cos\theta-\rho\sin\theta,\quad \frac{\partial y}{\partial\theta}=\frac{\partial\rho}{\partial\theta}\sin\theta+\rho\cos\theta,\quad \frac{\partial z}{\partial\theta}=0,$$

$$\frac{\partial x}{\partial z}=\frac{\partial \rho}{\partial z}\cos\theta,\quad \frac{\partial y}{\partial z}=\frac{\partial \rho}{\partial z}\sin\theta,\quad \frac{\partial z}{\partial z}=1$$

得知在曲面上的点 $(\theta,z)$ 的法线方向是

$$\left(\frac{\partial \rho}{\partial \theta}\sin\theta+\rho\cos\theta,\ -\frac{\partial \rho}{\partial \theta}\cos\theta+\rho\sin\theta,\ -\rho\frac{\partial \rho}{\partial z}\right).$$

由 (18) 可知它与 $z$ 轴的交角 $\alpha$(即点 $(\theta,z)$ 的倾角) 的余弦等于

$$\cos\alpha=\frac{-\rho\dfrac{\partial \rho}{\partial z}}{\sqrt{\left(\rho\dfrac{\partial \rho}{\partial z}\right)^2+\left(\dfrac{\partial \rho}{\partial \theta}\right)^2+\rho^2}}=\frac{1}{\sqrt{1+c^2}}$$

是一常数. 也就是说, 这曲面的切平面与地平面 (即 $xy$ 平面) 成一固定角度 $\alpha$. 我们来说明这样的曲面的几何性质.

从制高点向 $xy$ 平面作任一垂直平面, 这平面与该曲面的交线有次之性质. 这曲线上每一点的切线与 $xy$ 平面的交角为 $\alpha$. 因此, 它是一条直线.

从任一平面封闭曲线 $(l_0)$ 作底盘, 以任一投影在盘内的点 $(l_n)$ 作制高点. 通过制高点与底盘垂直的直线称为轴. 通过 $(l_0)$ 上任一点 $A$ 作一直线, 它在 $A$ 与轴所成的平面上, 与底盘的交角是 $\alpha$. 这样直线所成的图形便是适合 $\mathrm{B}_0=S$ 的图形.

所以, 如果有最高峰, 而且向下看没有陡峭的角度, 则仅有以下的曲面才能 $\mathrm{B}_0=S$. 底盘是圆或圆的若干切线形成的多角形或一些圆弧及一些切线所形成的图形, 轴的尖端在通过圆心而垂直于底盘的直线上 (见图 10).

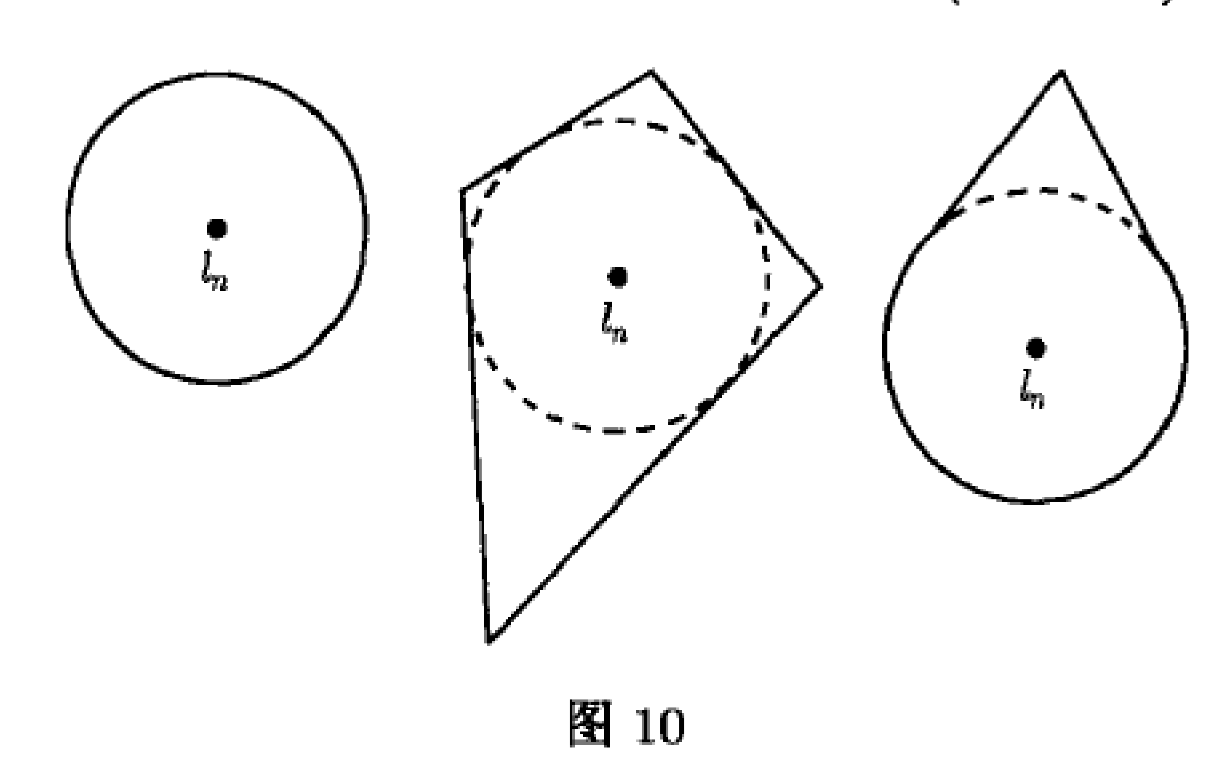



图 10

通俗些说, 只有蒙古包, 金字塔和一些由此复合出来的图形, 才能由 Волков 方法来无限逼近.

但什么时候Бa= $S$ 呢? 当然 $\mathrm{B}_0=S$ 的时候Бa= $S$ 除掉上面所求的曲面, 还有其他曲面否? 等等. 有! 证明如下, 从

$$\text{Ба}=\int_0^h\left|\int_0^{2\pi}f(z,\theta)d\theta\right|dz=\int_0^h\int_0^{2\pi}|f(z,\theta)|d\theta\ dz=S$$

由

$$\int_0^h \left( \int_0^{2\pi} |f(z,\theta)| d\theta - \left| \int_0^{2\pi} f(z,\theta) d\theta \right| \right) dz = 0.$$

因为积分号下的函数是非负的. 因此对任一 $z$ 常有

$$\int_0^{2\pi} |f(z,\theta)| d\theta = \left| \int_0^{2\pi} f(z,\theta) d\theta \right|$$

因此当固定 $z$ 时, $f(z,\theta)$ 的虚实部分之比是常数, 即方程 (18) 中的 $c$ 是仅为 $z$ 的函数. 所以仅有下面的曲面才能Бa= $S$. 高程相同之处, 曲面有相同的倾角. 用通俗的话说, 只有葫芦, 白塔 (北海), 才能由Бауман 方法来无限逼近.

现在我们来估计一下这两个方法给出的结果的偏差情况. 假定曲面上点的倾角的余弦介于两正常数 $\xi$ 与 $\eta$ 之间, 即

$$\xi \leqslant \cos\alpha \leqslant \eta,$$

即

$$\xi \leqslant \frac{-\rho\dfrac{\partial\rho}{\partial z}}{\sqrt{\left(\rho\dfrac{\partial\rho}{\partial z}\right)^2 + \left(\dfrac{\partial\rho}{\partial\theta}\right)^2 + \rho^2}} \leqslant \eta,$$

由此可得

$$\frac{\rho^2 + \left(\dfrac{\partial\rho}{\partial\theta}\right)^2}{\left(\rho\dfrac{\partial\rho}{\partial z}\right)^2 + \left(\dfrac{\partial\rho}{\partial\theta}\right)^2 + \rho^2} \geqslant 1 - \eta^2,$$

因而

$$\begin{aligned}
&\int_0^{2\pi}\int_0^h \sqrt{\rho^2 + \left(\frac{\partial\rho}{\partial\theta}\right)^2}\, dz\, d\theta \\
&\quad \geqslant \sqrt{1-\eta^2}\int_0^{2\pi} d\theta \int_0^h \sqrt{\left(\rho\frac{\partial\rho}{\partial z}\right)^2 + \left(\frac{\partial\rho}{\partial\theta}\right)^2 + \rho^2}\, dz \\
&\quad = \sqrt{1-\eta^2}S. \\
&\int_0^{2\pi}\int_0^h -\rho\frac{\partial\rho}{\partial z} dz\, d\theta \geqslant \xi S
\end{aligned}$$

因此

$$\text{Б0} \geqslant \sqrt{\xi^2 S^2 + (1-\eta^2)}S = \sqrt{1+\xi^2-\eta^2}S.$$

又因为 $1 > \eta \geqslant \xi > 0$, 所以

$$\frac{\xi}{\eta} \leqslant \sqrt{1+\xi^2-\eta^2}$$

(将两端平方, 此式即 $(\eta^2-\xi^2)(1-\eta^2) \geqslant 0$) 即得

$$\text{B0} \geqslant \frac{\xi}{\eta} S.$$

总而言之, 我们证明了下面的定理.

**定理 3** 若曲面 $\rho = \rho(z,\theta)(0 \leqslant z \leqslant h,\ 0 \leqslant \theta \leqslant 2\pi)$ 上任一点的倾角 $\alpha$ 的余弦都满足 $0 < \xi \leqslant \cos\alpha \leqslant \eta$, 则不等式

$$\frac{\xi}{\eta} S \leqslant \text{B0} \leqslant \text{Б}a \leqslant S \tag{19}$$

成立. B0$= S$ 的充要条件是曲面的任意点都有相同的倾角, Бa$= S$ 的充要条件是曲面在高程相等处的点有相同的倾角.

**6. 算法建议**

由定理 3 可以看出只有当曲面上的点的倾角变化不大时, Волков 方法才能得到精确结果, 而只有当曲面在相邻两高程间的点的倾角相差不大时, Бауман 方法才能给出精密的结果, 然而在其他情况下, 用这种方法的误差就可能比较大了.

因此我们建议如下的算法: 在等高线图上 (图 11), 通过制高点 $l_n$引进若干条放射线 $\theta_0,\ \theta_1,\ \cdots,\ \theta_{m-1}$, 其中 $\theta_j$ 的幅角等于 $\dfrac{2\pi j}{m}$. 放射线 $\theta_j, \theta_{j+1}$ 与等高线 $l_i, l_{i+1}$ 所围成的面积记为 $d_{ij}$; $l_i$ 被 $\theta_j,\ \theta_{j+1}$ 所截取的一段的长度记之为 $l_{ij}$.

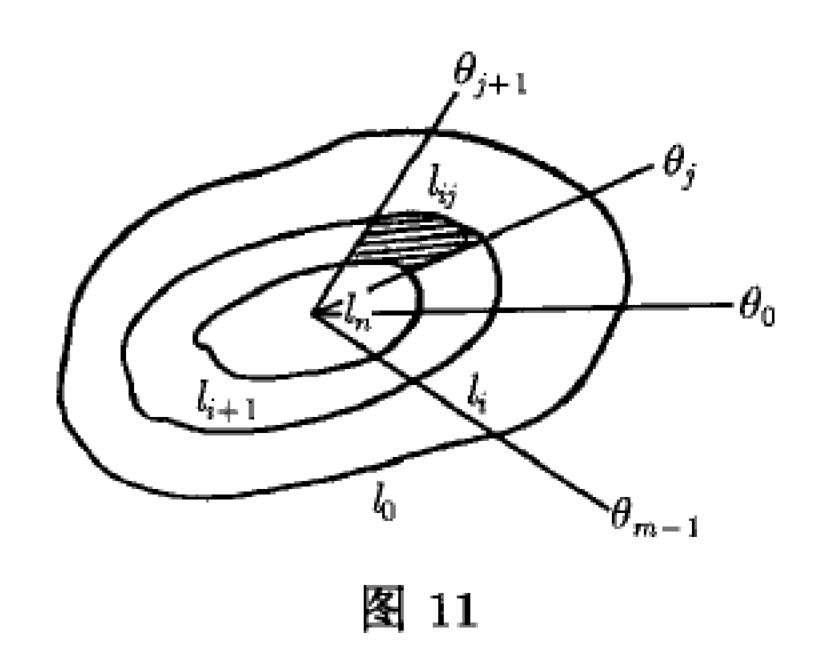



图 11

方法 I. a. $D_j = \sum\limits_{i=0}^{n-1} d_{ij}$ (等高线图在放射线 $\theta_j$ 与 $\theta_{j+1}$ 之间的面积);

b. $E_j = \left(\sum\limits_{i=0}^{n-1} l_{ij}\right) \Delta h$ (中间隔板在两直立墙壁之间的面积之和);

c. $\sigma_1 = \sum\limits_{j=0}^{m-1} \sqrt{D_j^2 + E_j^2}$ 就是所求曲面面积的渐近值.

方法 II. a. $e_{ij}=l_{ij}\Delta h$ (中间隔板在两直立墙壁之间的面积);

b. $\sigma_2=\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\sqrt{d_{ij}^2+e_{ij}^2}$ 就是所求曲面面积的渐近值.

与上段相同的方法可知

$$K=\int_0^{2\pi}\sqrt{\left(\int_0^h -\rho\frac{\partial\rho}{\partial z}dz\right)^2+\left(\int_0^h\sqrt{\rho^2+\left(\frac{\partial\rho}{\partial\theta}\right)^2}dz\right)^2}d\theta$$
$$=\int_0^{2\pi}\left|\int_0^h f(z,\theta)dz\right|d\theta \tag{20}$$

及

$$S=\int_0^{2\pi}\int_0^h\sqrt{\rho^2+\left(\frac{\partial\rho}{\partial\theta}\right)^2+\left(\rho\frac{\partial\rho}{\partial z}\right)^2}dz\ d\theta$$
$$=\int_0^{2\pi}\int_0^h|f(z,\theta)|dz\ d\theta \tag{21}$$

分别为当 $n\to\infty$, $m\to\infty$ 时, $\sigma_1$ 与 $\sigma_2$ 所趋近的数 (关于 $f(z,\theta)$ 的定义请参看 (12) 式).

显然 $\mathrm{B}_0\leqslant K\leqslant S$(见 (10)), 同上段的方法可知 $K=S$ 的充要条件为曲面为直纹面. 由于 $\sigma_2$ 趋于真面积, 所以方法 II 最为精密可靠.

## 参 考 文 献

[1] В. И. Бауман, К вопросу оподсчета запасов полезных ископаемых, Горный журнал. Декабрь, 1908.

[2] 乌沙阔夫 (И. Н. Ушаков), 矿藏几何学, 煤炭工业出版社, 1957.

[3] 雷若夫 (П. А. Рыжов), 矿体几何学, 地质出版社, 1957.

[4] 依札克松 (С. С. Изаксон), 矿产储量计算的验算和计算误差的确定, 煤炭工业出版社, 1958.

[5] 伏尔科夫 (Н. М. Волков), 量图原理和方法, 1950.

[6] 陆漱芬, 在等高线地图上计算地表面面积的问题, 测量制量学报, 第 4 卷第 1 期, 1960.

(原载 1961 年第 1 期"数学学报")

# 谈谈与蜂房结构有关的数学问题

人类识自然,
探索穹研,
花明柳暗别有天,
谲诡神奇满目是,
气象万千.
往事几百年,
祖述前贤,
瑕疵讹谬犹盈篇,
蜂房秘奥未全揭,
待咱向前.

## 楔　　子

先谈谈我接触到和思考这问题的过程. 始之以“有趣”. 在看到了通俗读物上所描述的自然界的奇迹之一 —— 蜂房结构的时候, 觉得趣味盎然, 引人入胜. 但继之而来的却是“困惑”. 中学程度的读物上所提出的数学问题我竟不会, 或说得更确切些, 我竟不能在脑海中想象出一个几何模型来, 当然我更不能列出所对应的数学问题来了, 更不要说用数学方法来解决这个问题了! 在列不出数学问题, 想象不出几何模型的时候, 咋办? 感性知识不够, 于是乎请教实物, 找个蜂房来看看. 看了之后, 了解了, 原来如此, 问题形成了, 因而很快地初步解决了. 但解法中用了些微积分, 因而提出一个问题, 能不能不用微积分, 想出些使中学同学能懂的初等解法. 这样就出现了本文的第五节“浅化”(在这段中还将包括南京师范学院附中老师和同学给我提出的几种不同解法. 这种听了报告就动手动脑的风气是值得称道的). 问题解得是否全面? 更全面地考虑后, 引出一个“难题”. 这难题的解决需要些较高深或较繁复的数学. 在本文中我作了些对比, 以便看出蜂房的特点来.

在深入探讨一下之后发现, 容积一样而用材最省的尺寸比例竟不是实测下来的数据, 因而使我们怀疑前人已得的结论, 因而发现问题的提法也必须改变, 似乎应当是: 以蜜蜂的身长腰围为准, 怎样的蜂房才最省材料. 这样问题就更进了一步, 不是仅仅乎依赖于空间形式与数量关系的数学问题了, 而是与生物体统一在一起的问题了, 这问题的解答, 不是本书的水平所能胜任的.

问题看清了，解答找到了．但还不能就此作结，随之而来的是浮想联翩．更丰富更多的问题，在这小册子上是写不完的，并且不少已经超出了中学生水平．但在最后我还是约略地提一下，写了几节中学生可能看不懂的东西，留些咀嚼余味罢！

总之，我做了一个习题．我把做习题的源源本本写下来供中学同学参考，请读者指正．

## 一 有 趣

我把我所接触到的通俗读物中有关蜂房的材料摘引几条（有些用括号标出的问句或问号是作者添上的）．

如果把蜜蜂大小放大为人体的大小，蜂箱就会成为一个悬挂在几乎达 20 公顷的天顶上的密集的立体市镇．

一道微弱的光线从市镇的一边射来，人们看到由高到低悬挂着一排排一列列五十层的建筑物．

耸立在左右两条街中间的高楼上，排列着薄墙围成的既深又矮的，成千上万个六角形巢房．

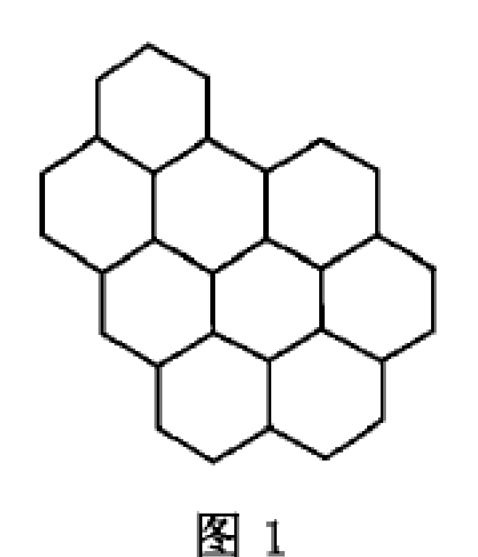


图 1

为什么是六角形？这到底有什么好处？十八世纪初，法国学者马拉尔琪曾经测量过蜂窝的尺寸，得到一个有趣的发现，那就是六角形窝洞的六个角，都有一致的规律：钝角等于 109° 28′，锐角等于 70° 32′．（对吗？）

难道这是偶然的现象吗？法国物理学家列奥缪拉由此得到一个启示，蜂窝的形状是不是为了使材料最节省而容积最大呢？（确切的提法应当是，同样大的容积，建筑用材最省；或同样多的建筑材料，造成最大容积的容器．）

列奥缪拉去请教巴黎科学院院士瑞士数学家克尼格．他计算的结果，使人非常震惊．因为他从理论上的计算，要消耗最少的材料，制成最大的菱形容器（？），它的角度应该是 109° 26′ 和 70° 34′．这与蜂窝的角度仅差 2 分．

后来，苏格兰数学家马克劳林又重新计算了一次，得出的结果竟和蜂窝的角度完全一样．后来发现，原来是克尼格计算时所用的对数表（？）印错了！

小小蜜蜂在人类有史以前所已经解决的问题，竟要十八世纪的数学家用高等数

学才能解决呢!

这些是多么有趣的描述呀! “小小蜜蜂”, “科学院院士”,“高等数学”, “对数表印错了”! 真是引人入胜的描述呀! 启发人们思考的描述呀!

诚如达尔文说得好:“巢房的精巧构造十分符合需要, 如果一个人看到巢房而不备加赞扬, 那他一定是个糊涂虫.” 自然界的奇迹如此, 人类认识这问题的过程又如此, 怎能不引人入胜呢!

## 二　困　惑

是的, 真有趣. 这个十八世纪数学家所已经解决的问题, 我们会不会? 如果会, 要用怎样的高等数学? 大学教授能不能解? 大学高年级学生能不能解? 我们现在是二十世纪了, 大学低年级学生能不能解? 中学生能不能解? 且慢! 这到底是个什么数学问题? 什么样的六角形窝洞的钝角等于 $109^\circ 28'$, 锐角等于 $70^\circ 32'$? 不懂! 六角形六内角的和等于 $(6-2)\pi=4\pi=720^\circ$, 每个角平均 $120^\circ$, 而 $109^\circ 28'$, 与 $70^\circ 32'$ 都小于 $120^\circ$, 因而不可能有这样的六角形.

既说“蜂窝是六角形的”, 又说“它是菱形容器”, 所描述的到底是个什么样子? 六角形和菱形都是平面图形的术语, 怎样用来刻画一个立体结构? 不懂!

困恼! 不要说解问题了, 连个蜂窝模型都摸不清. 问题钉在心上了! 这样想, 那样推, 无法在脑海形成一个形象来. 设想出了几个结构, 算来算去, 都与事实不符, 找不出这样的角度来. 这还不只是数学问题, 而必须请教一下实物, 看看蜂房到底是怎样的几何形状, 所谓的角到底是指的什么角!

## 三　访　实

解除困恼的最简单的办法是撤退. 是的, 我们有一千个理由可以撤退, 像这是已经解决了的问题呀! 这不是属于我们研究的范围内的问题呀! 这还不是确切的数学问题呀! 这些理由中只要有一个抬头, 我们就将失去了一个锻炼的机会. 一千个理由顶不上一个理由, 就是不会! 不会就得想, 就得想到水落石出来. 空间的几何图形既然还属茫然, 当然就必须请教实物. 感谢昆虫学家刘崇乐教授, 他给了我一个蜂房, 使我摆脱了困境.

画一支铅笔怎样画? 是否把它画成为如图 2 那样?

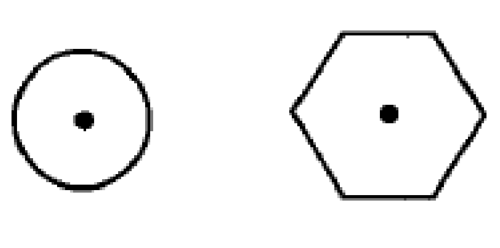


图 2

有人说这不像, 我说很像. 我是从近处正对着铅笔头画的. 这是写实, 但是并不足以刻画出铅笔的形态来. 我们的图 1(如第一节的说明) 就是用“正对铅笔头的方法”画出来的, 当然没有了立体感, 更无法显示出蜂房内部的构造情况.

看到了实物, 才知道既说“六角”又说“菱形”的意义. 原来是' 正面看来, 蜂房是由一些正六边形所组成的. 既然是正六边形, 那就每一角都是 120°, 并没有什么角度的问题. 问题在于房底. 蜂房并非六棱柱, 它的底部都是由三个菱形所拼成的. 图 3 是蜂房的立体图. 这个图比较清楚些, 但还是得用各种分图及说明来解释清楚. 说得更具体些, 拿一支六棱柱的铅笔, 未削之前, 铅笔一端的形状是正六角形 $ABCDEF$(图 4). 通过 $AC$, 一刀切下一角, 把三角形 $ABC$ 搬置 $AP'C$ 处; 过 $AE, CE$ 切如此同样三刀, 所堆成的形状就如图 5 那样, 而蜂巢就是由两排这样的蜂房底部和底部相接而成的.

因而初步形成了以下的数学问题了:

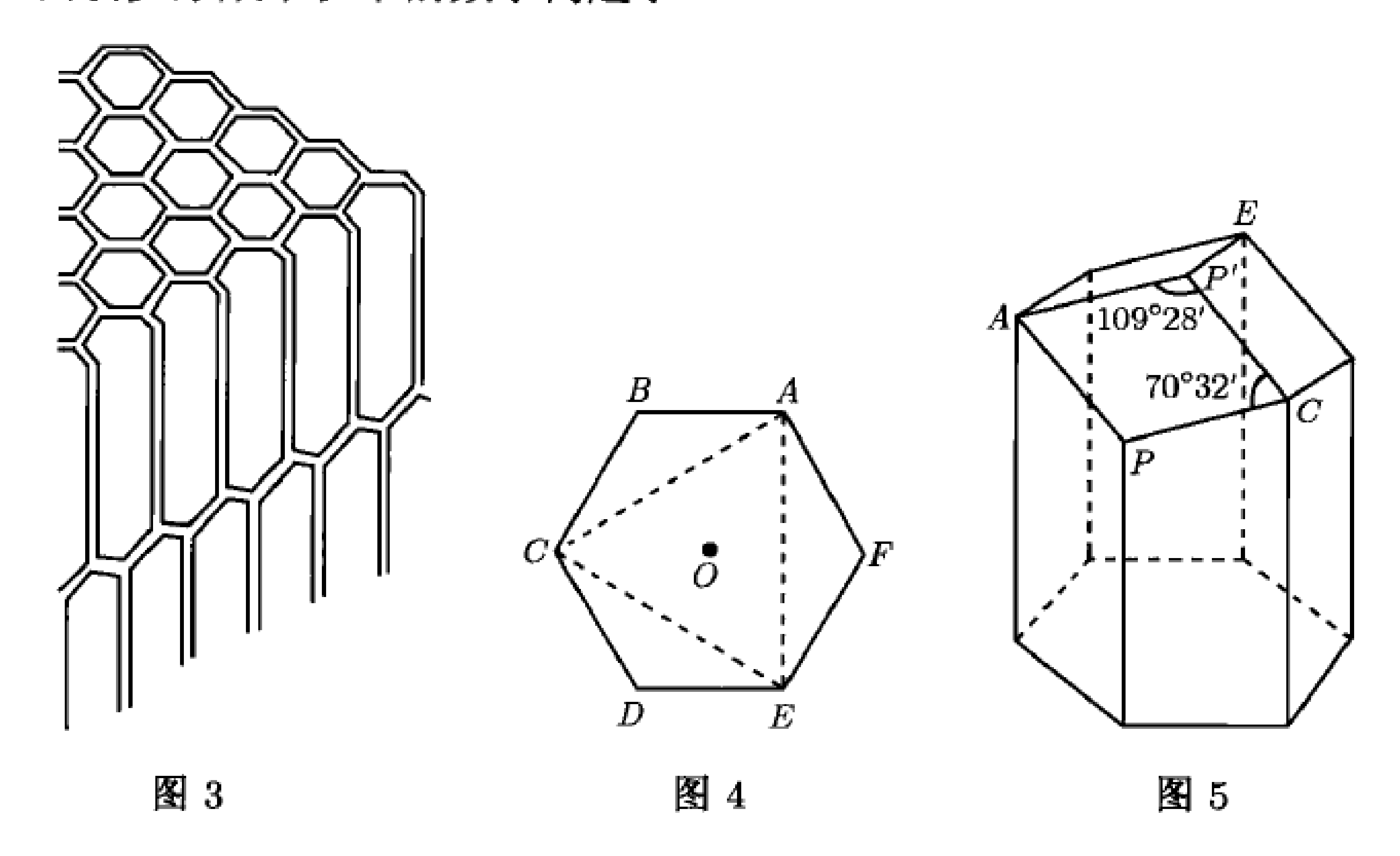



图 3　　图 4　　图 5

怎样切出来使所拼成的三个菱形做底的六面柱的表面积最小?

为什么说是“初步”? 且待第六、第七节分解. 下节中首先解决这个简单问题.

(读者试利用这机会来考验一下自己对几何图形的空间想象能力. 这样的图形可以排成密切无间的蜂窝.)

## 四　解　　题

假定六棱柱的边长是 1, 先求 $AC$ 的长度. $ABC$ 是腰长为 1, 夹角为 120° 的等腰三角形. 以 $AC$ 为对称轴作一个三角形 $AB'C$(图 6). 三角形 $ABB'$ 是等边三角

形. 因此,

$$\frac{1}{2}AC=\sqrt{1-\left(\frac{1}{2}\right)^2}=\frac{\sqrt{3}}{2},$$

即得 $AC=\sqrt{3}$.

把图 5 的表面分成六份, 把其中之一摊平下来, 得出图 7 的形状. 从一个宽为 1 的长方形切去一角, 切割处成边 $AP$. 以 $AP$ 为腰, $\frac{\sqrt{3}}{2}$ 为高作等腰三角形. 问题: 怎样切才能使所作出的图形的面积最小?

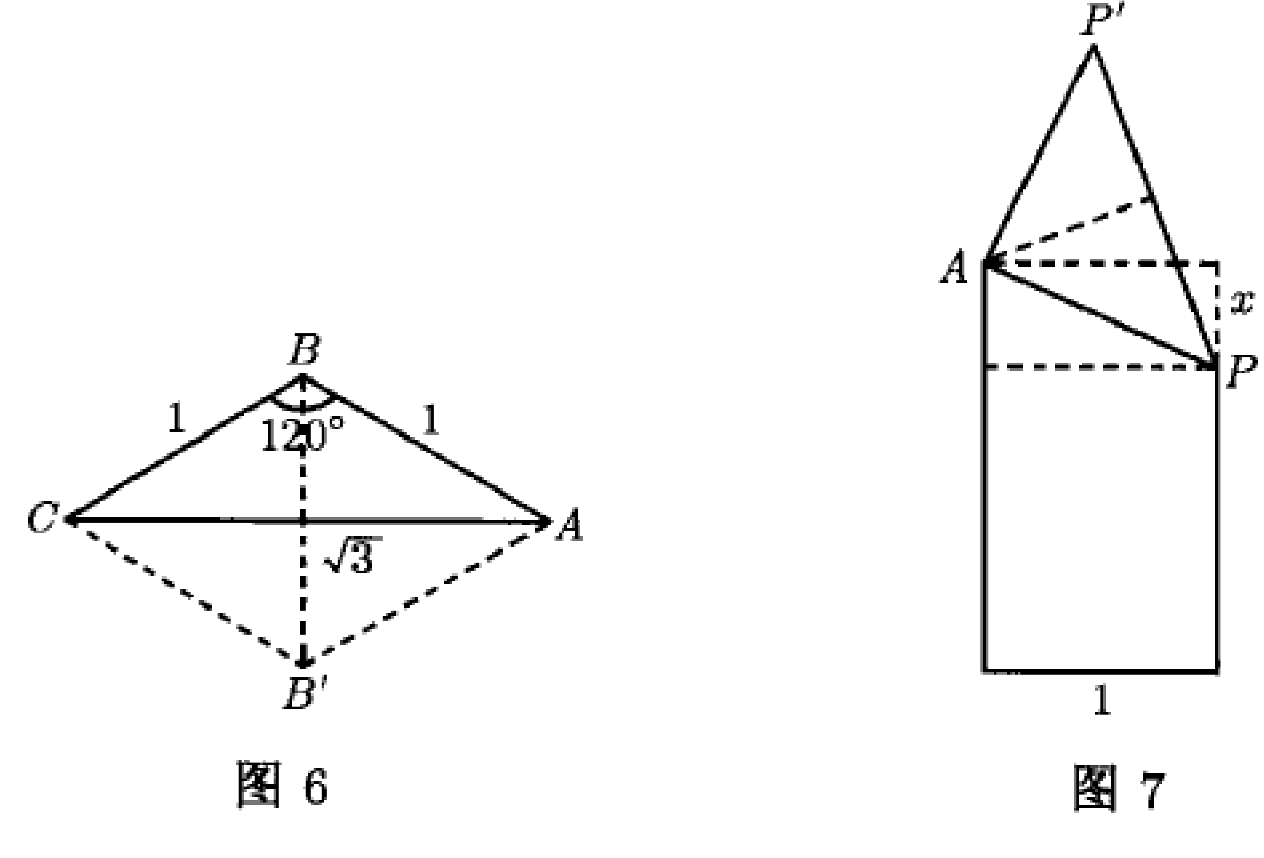



图 6　　图 7

假定被切去的三角形的高是 $x$. 从矩形中所切去的面积等于 $\frac{1}{2}x$. 现在看所添上的三角形 $APP'$ 的面积. $AP$ 的长度是 $\sqrt{1+x^2}$, 因此 $PP'$ 的长度等于

$$2\sqrt{(1+x^2)-\frac{3}{4}}=\sqrt{1+4x^2},$$

因而三角形 $APP'$ 的面积等于

$$\frac{\sqrt{3}}{4}\sqrt{1+4x^2}.$$

问题再变而为求

$$-\frac{1}{2}x+\frac{\sqrt{3}}{4}\sqrt{1+4x^2}$$

的最小值的问题.

念过微积分的读者立刻可以用以下的方法求解: (没有学过微积分的读者可以略去以下这一段.)

求

$$f(x)=-\frac{1}{2}x+\frac{\sqrt{3}}{4}\sqrt{1+4x^2}$$

的微商, 得

$$f'(x)=-\frac{1}{2}+\frac{\sqrt{3}x}{\sqrt{1+4x^2}}.$$

由 $f'(x)=0$, 解得 $1+4x^2=12x^2$, $x=\frac{1}{\sqrt{8}}$. 又

$$f''(x)=\frac{\sqrt{3}}{\sqrt{1+4x^2}}-\frac{4\sqrt{3}x^2}{(1+4x^2)^{3/2}}=\frac{\sqrt{3}}{(1+4x^2)^{3/2}}>0,$$

因而当 $x=\frac{1}{\sqrt{8}}$ 时给出极小值

$$f\left(\frac{1}{\sqrt{8}}\right)=-\frac{1}{4\sqrt{2}}+\frac{\sqrt{3}}{4}\times\frac{\sqrt{3}}{\sqrt{2}}=\frac{1}{\sqrt{8}}.$$

这一节说明了当 $x=\frac{1}{\sqrt{8}}$ 时取最小值, 即在一棱上过 $x=\frac{1}{\sqrt{8}}$ 处 (图 5 中 $P$ 点) 以及与该棱相邻的二棱的端点 (图 5 中 $A,C$ 点) 切下来拼上去的图形的表面积最小.

用 $\gamma$ 表示三角形 $APP'$ 两腰的夹角 $\angle PAP'$. $\gamma$ 的余弦由以下的余弦公式给出:

$$2(1+x^2)\cos\gamma=2(1+x^2)-(1+4x^2)=1-2x^2,$$

即

$$\cos\gamma=\frac{1-2x^2}{2(1+x^2)}=\frac{3}{8}\Big/\left(1+\frac{1}{8}\right)=\frac{1}{3}.$$

因此得出 $\gamma=70°\ 32'$.

把问题说得更一般些, 以边长为 $a$ 的正方边形为底, 以 $b$ 为高的六棱柱, 其六个顶点顺次以 $ABCDEF$ 标出 (图 8). 过 $B$(或 $D$ 或 $F$) 棱距顶点为 $\frac{1}{\sqrt{8}}a$ 处及 $A,C$(或 $C,E$ 或 $E,A$) 作一平面; 切下三个四面体, 反过来堆在顶上, 得一以三个菱形做底的六棱尖顶柱. 现在算出这六棱尖顶柱的体积和表面积:

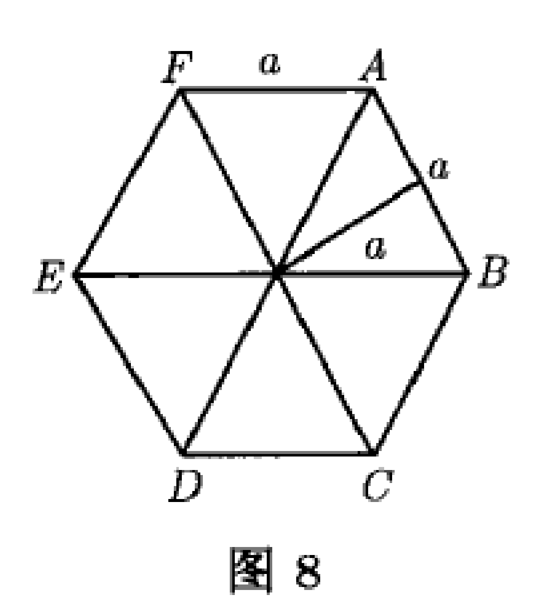



图 8

体积等于以边长为 $a$ 的正六角形的面积乘高 $b$. 即

$$6\times\frac{1}{2}a\times\frac{\sqrt{3}}{2}a=\frac{3\sqrt{3}}{2}a^2$$

乘以 $b$, 即得

$$\frac{3\sqrt{3}}{2}a^2b.$$

表面积等于六棱柱的侧面积 $6ab$ 加上六倍的 $\frac{1}{\sqrt{8}}a^2\left[\text{也就是}f\left(\frac{1}{\sqrt{8}}\right)a^2=\frac{1}{\sqrt{8}}a^2\right]$, 即

$$6ab+\frac{6}{\sqrt{8}}a^2=6a\left(b+\frac{a}{\sqrt{8}}\right).$$

## 五 浅 化

没有读过微积分的读者不要着急. 在我解决了这问题之后, 当然就想到了要不要用微积分, 能不能找到一个中学生所能理解的解法. 有的, 而且很不少.

**方法一** 我们需要用以下的结果 (或称为算术中项大于几何中项). 当 $a\geqslant 0$, $b\geqslant 0$ 时, 常有

$$\frac{1}{2}(a+b)\geqslant\sqrt{ab}, \tag{1}$$

当 $a=b$ 时取等号, 当 $a\neq b$ 时到不等号. 这一结论可由不等式

$$(\sqrt{a}-\sqrt{b})^2\geqslant 0 \tag{2}$$

立刻推出.

现在试来解决问题. 命 $2x=t-\frac{1}{4t}(t>0)$, 则

$$\begin{aligned}f(x)&=-\frac{1}{2}x+\frac{\sqrt{3}}{4}\sqrt{1+4x^2}\\&=-\frac{1}{4}\left(t-\frac{1}{4t}\right)+\frac{\sqrt{3}}{4}\left(t+\frac{1}{4t}\right)\\&=\frac{\sqrt{3}-1}{4}t+\frac{\sqrt{3}+1}{4}\times\frac{1}{4t}.\end{aligned}$$

由 (1) 得出

$$f(x)\geqslant 2\sqrt{\frac{(\sqrt{3}-1)(\sqrt{3}+1)}{4^3}}=\frac{1}{\sqrt{8}}.$$

并且知道仅当

$$\frac{\sqrt{3}-1}{4}t=\frac{\sqrt{3}+1}{4}\times\frac{1}{4t}$$

时取等号. 即, 当

$$4t^2=\frac{\sqrt{3}+1}{\sqrt{3}-1}=\frac{(1+\sqrt{3})^2}{2},\ t=\frac{1+\sqrt{3}}{2\sqrt{2}};$$

而当

$$x = \frac{1}{2}\left[\frac{1+\sqrt{3}}{2\sqrt{2}} - \frac{2\sqrt{2}}{4(1+\sqrt{3})}\right]$$
$$= \frac{1}{2}\left(\frac{1+\sqrt{3}}{2\sqrt{2}} + \frac{1-\sqrt{3}}{2\sqrt{2}}\right) = \frac{1}{\sqrt{8}}$$

时,$f(x)$ 取得小值 $\frac{1}{\sqrt{8}}$.

**方法二** 在式子

$$[\lambda(\sqrt{1+4x^2}+2x)^{\frac{1}{2}} - \mu(\sqrt{1+4x^2}-2x)^{\frac{1}{2}}]^2$$
$$= 2(\lambda^2-\mu^2)x + (\lambda^2+\mu^2)\sqrt{1+4x^2} - 2\lambda\mu \geqslant 0$$

中, 取 $2(\lambda^2-\mu^2) = -\frac{1}{2}$, $\lambda^2+\mu^2 = \frac{\sqrt{3}}{4}$, 即得

$$-\frac{1}{2}x + \frac{\sqrt{3}}{4}\sqrt{1+4x^2} \geqslant 2\lambda\mu = \sqrt{(\lambda^2+\mu^2)^2 - (\lambda^2-\mu^2)^2}$$
$$= \sqrt{\frac{3}{4^2} - \frac{1}{4^2}} = \frac{1}{\sqrt{8}}.$$

并且仅当 $\lambda^2(\sqrt{1+4x^2}+2x) = \mu^2(\sqrt{1+4x^2}-2x)$ 时取等号, 即

$$(\lambda^2-\mu^2)\sqrt{1+4x^2} + 2(\lambda^2+\mu^2)x$$
$$= -\frac{1}{4}\sqrt{1+4x^2} + \frac{\sqrt{3}}{2}x = 0$$

时取等号, 解得 $x = \frac{1}{\sqrt{8}}$.

**方法三** 命 $2x = \operatorname{tg}\theta$, 则

$$f(x) = \frac{-\frac{1}{4}\sin\theta + \frac{1}{4}\sqrt{3}}{\cos\theta} = \alpha \times \frac{1-\sin\theta}{\cos\theta} + \beta \times \frac{1+\sin\theta}{\cos\theta}$$
$$\geqslant 2\sqrt{\alpha\beta\frac{1-\sin^2\theta}{\cos^2\theta}} = 2\sqrt{\alpha\beta},$$

这儿$\alpha+\beta = \frac{\sqrt{3}}{4}$, $-\alpha+\beta = -\frac{1}{4}$, 不难由此解得答案.

方法虽是三个, 实质仅有一条, 转来转去仍然是依据了 $a^2+b^2-2ab = (b-a)^2 \geqslant 0$.

南京师范学院附中的老师和同学们又提供了以下的四个证明 (方法四至方法七).

**方法四**　令

$$y=-\frac{1}{2}x+\frac{\sqrt{3}}{4}\sqrt{1+4x^2},$$

故

$$y+\frac{1}{2}x=\frac{\sqrt{3}}{4}\sqrt{1+4x^2},$$

两边平方并加以整理得

$$x^2-2yx+\frac{3}{8}-2y^2=0. \tag{3}$$

因为 $x$ 为实数, 故二次方程 (3) 的判别式

$$\Delta=y^2-\frac{3}{8}+2y^2=3y^2-\frac{3}{8}\geqslant 0,$$

而 $y$ 必大于 0, 因此 $y$ 的最小值是 $\frac{1}{\sqrt{8}}$. 以此代入 (3), 则

$$x=\frac{1}{\sqrt{8}}.$$

**方法五**　设

$$\sqrt{1+4x^2}=2x+t,\ (t>0)$$

由此得

$$x=\frac{1-t^2}{4t},$$

因此

$$\sqrt{1+4x^2}=\frac{1-t^2}{2t}+t=\frac{1+t^2}{2t}.$$

故

$$\begin{aligned}f(x)=&\frac{t^2-1}{8t}+\frac{\sqrt{3}(t^2+1)}{8t}\\=&\frac{1}{8}\times\frac{(\sqrt{3}+1)t^2+(\sqrt{3}-1)}{t}\\=&\frac{1}{8}\left[(\sqrt{3}+1)t+(\sqrt{3}-1)\frac{1}{t}\right]\\\geqslant&\frac{1}{8}\times 2\sqrt{(\sqrt{3}+1)(\sqrt{3}-1)}=\frac{1}{\sqrt{8}}.\end{aligned}$$

由此不难解出问题.

**方法六**　设

$$2x=\operatorname{tg}\theta.$$

则

$$y=f(x)=\frac{\sqrt{3}}{4}\sec\theta-\frac{1}{4}\text{tg }\theta,$$

即

$$4y\cos\theta+\sin\theta=\sqrt{3},$$

$$\sqrt{1+(4y)^2}\sin(\theta+\varphi)=\sqrt{3},$$

这儿 $\varphi$ 由 $\text{tg }\varphi=4y$ 决定. 因此,

$$\sin(\theta+\varphi)=\sqrt{\frac{3}{1+16y^2}}\leqslant 1,$$

即

$$1+16y^2\geqslant 3,$$

故 $y$ 的最小值为 $\frac{1}{\sqrt{8}}$. 这时 $\text{tg }\varphi=\sqrt{2}$, $\text{ctg }\varphi=\frac{\sqrt{2}}{2}$, $\sin(\theta+\varphi)=1$. 因此

$$\theta+\varphi=2k\pi+\frac{\pi}{2}.\quad(k=0,1,\cdots)$$

于是 $\text{tg }\theta=\text{ctg }\varphi=\frac{\sqrt{2}}{2}$, $x=\frac{1}{2}\text{tg }\theta=\frac{1}{\sqrt{8}}$.

**方法七**

首先证明, 当 $b\geqslant 1, x\geqslant 0$ 时下列不等式成立:

$$\sqrt{b(1+x)}-\sqrt{x}\geqslant\sqrt{b-1};\tag{4}$$

且仅当 $x=\frac{1}{b-1}$ 时等号成立.

**证**

$$[(b-1)x-1]^2=(b-1)^2x^2-2(b-1)x+1\geqslant 0.$$

故

$$(b+1)^2x^2+2(b+1)x+1\geqslant 4bx(1+x)>0,$$

$$(b+1)x+1\geqslant 2\sqrt{b(x+1)}\times\sqrt{x},$$

$$b(x+1)-2\sqrt{b(x+1)}\times\sqrt{x}+x\geqslant b-1.$$

即

$$(\sqrt{b(x+1)}-\sqrt{x})^2\geqslant b-1>0.$$

则

$$\sqrt{b(x+1)}-\sqrt{x}\geqslant\sqrt{b-1}.$$

这样, 不等式 (4) 得证. 由此

$$\begin{aligned}&-\frac{1}{2}x+\frac{\sqrt{3}}{4}\sqrt{1+4x^2}\\&=\frac{1}{4}[\sqrt{3(1+4x^2)}-\sqrt{4x^2}]\geqslant\frac{1}{4}\times\sqrt{2};\end{aligned}$$

仅当 $4x^2=\frac{1}{2}$ 时 (此时 $b=3$) 等号成立, 即得问题之解.

**方法八**　(北京师范大学附属实验中学某高一同学的解法)

由

$$y=-\frac{1}{2}x+\frac{\sqrt{3}}{4}\sqrt{1+4x^2},$$

清理方根号得出

$$y^2+xy=\frac{3}{16}+\frac{1}{2}x^2,$$

即

$$y^2-\frac{1}{8}=\frac{1}{3}(x-y)^2.$$

可知当 $x=y=\frac{1}{\sqrt{8}}$ 时, $y$ 取最小值.

读者试分析这些证法的原则性的共同点或不同点 (例如: 配方).

## 六　慎　　微

我们必须小心在意, 不要以为前所提出的几何问题和我们上两节所讨论的代数问题是完全等价的了. 在几何问题中, 切割处不能超过六棱柱的高度, 也就是高度 $b$ 必须 $\geqslant\frac{1}{\sqrt{8}}a$ 才有意义. 如果 $b<\frac{1}{\sqrt{8}}a$, 应当怎样切才对? 是否就是通过上底的 $AC$ 及下底的 $B'$ 所切出的方法, 共切三刀所得出的图形 (图 9)?

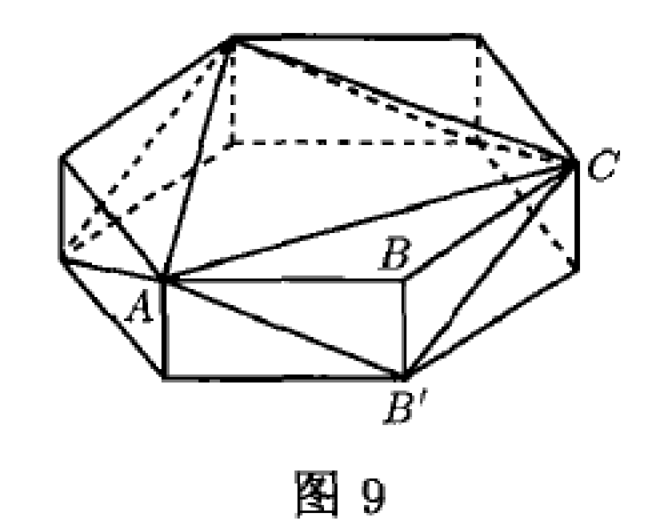



图 9

# 七　切　方

从以上的问题立刻可以联想到, 以六棱柱为基础, 还有没有其他的切拼方法? 例如, 不是尖顶六棱柱, 而是屋脊六棱柱行不行? 由四方柱出发行不行? 用四方柱怎样切下接上最好? 读者不妨多方设想. 我现在举以下二例:

1. 从边长为 1 的正四方柱的 $\frac{1}{4}$ 处切下一个三角柱堆到顶上, 对边也如此切, 也如此堆上去 (图 10, 参看图 14), 堆好之后得一方柱上加一屋脊的形状. 求切在何处, 表面积最小?

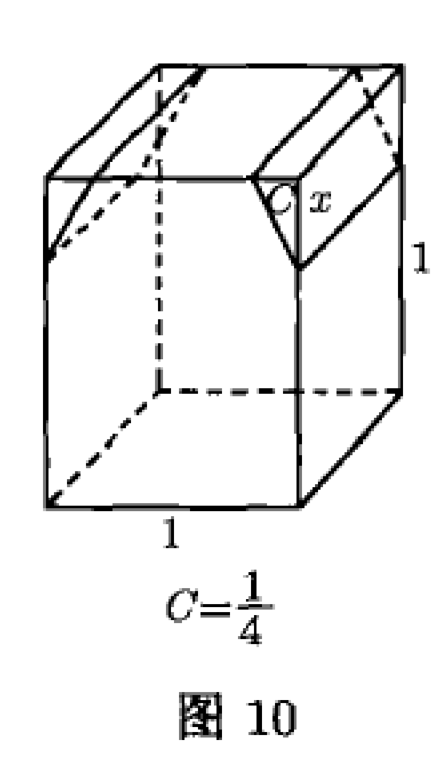



图 10

假定在棱上距顶点 $x$ 处切. 一刀使侧面少去一个矩形, 面积是 $x$(并且同时还少掉两个三角形, 但是把切下来的三角柱搬置顶上以后, 此两个三角形仍为柱体的侧面, 因此实际上并没有少), 添上三角柱翻开后暴露出的两个侧面. 其总面积是 $2\sqrt{x^2+\frac{1}{4^2}}$. 因此, 问题成为求 $-x+2\sqrt{x^2+\frac{1}{4^2}}$ 的最小值.

不难求出, 当 $x=\frac{1}{4\sqrt{3}}$ 时, 此面积取最小值

$$-\frac{1}{4\sqrt{3}}+2\sqrt{\frac{1}{4^2\times 3}+\frac{1}{4^2}}=-\frac{1}{4\sqrt{3}}+\frac{1}{\sqrt{3}}=\frac{\sqrt{3}}{4}.$$

2. 如果把"切边"改为"切角", 即过两边中点及棱上距顶占为 $x$ 处切下四面体堆上去的情况 (图 11).

一刀切去侧面两个三角形, 其总面积为 $2\times\frac{1}{2}x\times\frac{1}{2}=\frac{x}{2}$; 添上两个边长为

$$\sqrt{x^2+\frac{1}{2^2}},\ \sqrt{x^2+\frac{1}{2^2}},\ \sqrt{\frac{1}{2^2}+\frac{1}{2^2}}$$

的三角形 (图 12), 其总面积是

$$2\times\frac{1}{2}\times\frac{1}{\sqrt{2}}\times\sqrt{x^2+\frac{1}{2^2}-\left(\frac{1}{2\sqrt{2}}\right)^2}=\frac{1}{\sqrt{2}}\times\sqrt{x^2+\frac{1}{8}}.$$

问题成为求

$$-\frac{x}{2}+\frac{1}{\sqrt{2}}\times\sqrt{x^2+\frac{1}{8}}$$

的最小值.

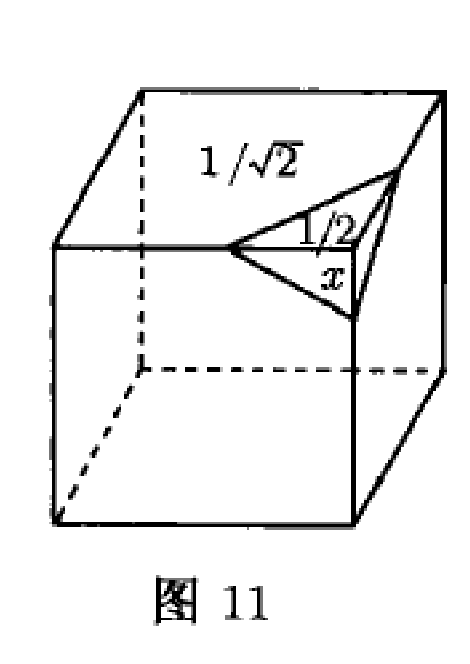



图 11

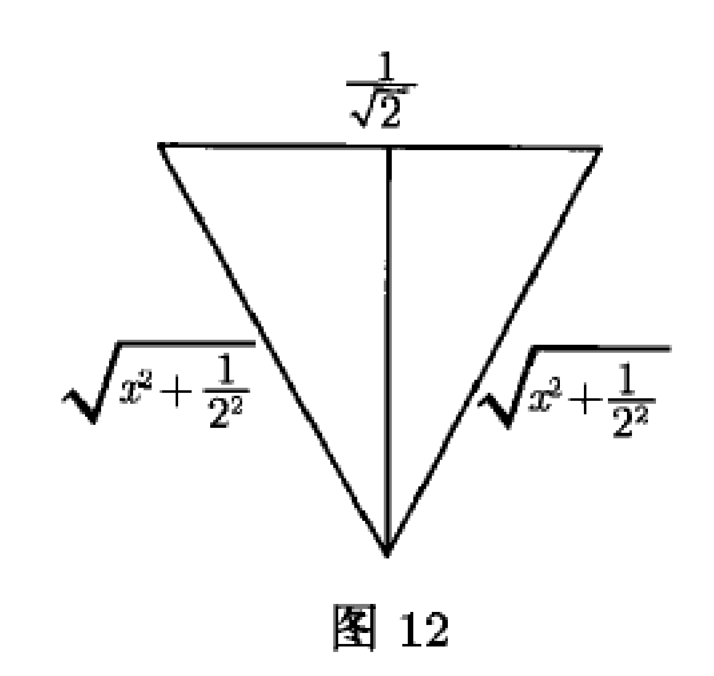



图 12

不难求出当 $x=\frac{1}{\sqrt{8}}$ 时, 即得最小值 $\frac{1}{2\sqrt{8}}$.

两种切法相比, 前一法添上二块大小是 $\frac{\sqrt{3}}{4}$ 的面积, 后一法添上四块大小是 $\frac{1}{2\sqrt{8}}$ 的面积. 由于

$$4\times\frac{1}{2\sqrt{8}}=\frac{\sqrt{2}}{2}<2\times\frac{\sqrt{3}}{4}=\frac{\sqrt{3}}{2},$$

所以第二种切法更好些.

把第一种切法讲得更一般些: 四方柱的底是边长为 $a$ 的正方形, 高是 $b$. 从四方形边的 $\frac{1}{4}a$ 处及棱上 $\frac{1}{4\sqrt{3}}a$ 处切下一个三角柱, 堆到顶上. 则所得屋脊四方柱的体积仍为 $a^2b$, 而表面积 (不算底面) 为

$$4ab+2\times\frac{\sqrt{3}}{4}a^2=4a\left(b+\frac{\sqrt{3}}{8}a\right).$$

第二种切法的一般情况则是: 四方柱的底是边长为 $a$ 的正方形, 高为 $b$. 从四方形两邻边的中点及棱上 $\frac{1}{\sqrt{8}}a$ 处切下四个四面体, 堆到顶上形成一个尖顶四方柱, 其体积仍是 $a^2b$, 而表面积 (不算底面) 是 $4ab+\frac{1}{\sqrt{2}}a^2$.

## 八　疑　　古

以上虽然讲了不少, 我们还没有回答出 “同样的体积, 哪一种模型需要建筑材料最少” 的问题. 在处理这问题之前, 先证明以下的不等式

如果 $a\geqslant 0, b\geqslant 0, c\geqslant 0$, 则

$$\frac{1}{3}(a+b+c)\geqslant(abc)^{\frac{1}{3}},\tag{1}$$

且仅当 $a=b=c$ 时取等号.

其证明可由以下的恒等式推出：

$$\begin{aligned}&a+b+c-3(abc)^{\frac{1}{3}}\\&=(a^{\frac{1}{3}}+b^{\frac{1}{3}}+c^{\frac{1}{3}})[a^{\frac{2}{3}}+b^{\frac{2}{3}}+c^{\frac{2}{3}}-(ab)^{\frac{1}{3}}-(bc)^{\frac{1}{3}}-(ca)^{\frac{1}{3}}]\\&=\frac{1}{2}(a^{\frac{1}{3}}+b^{\frac{1}{3}}+c^{\frac{1}{3}})[(a^{\frac{1}{3}}-b^{\frac{1}{3}})^2+(b^{\frac{1}{3}}-c^{\frac{1}{3}})^2+(c^{\frac{1}{3}}+a^{\frac{1}{3}})^2].\end{aligned}$$

**定理 1** 体积为 $V$ 的尖顶六棱柱的表面积 (不算底面) 的最小值是 $3\sqrt{2}V^{\frac{2}{3}}$, 而且仅当六角形边长是 $\sqrt{\frac{2}{3}}V^{\frac{1}{3}}$, 高度是 $\frac{1}{2}\sqrt{3}V^{\frac{1}{3}}$ 时取这最小值 (图 13).

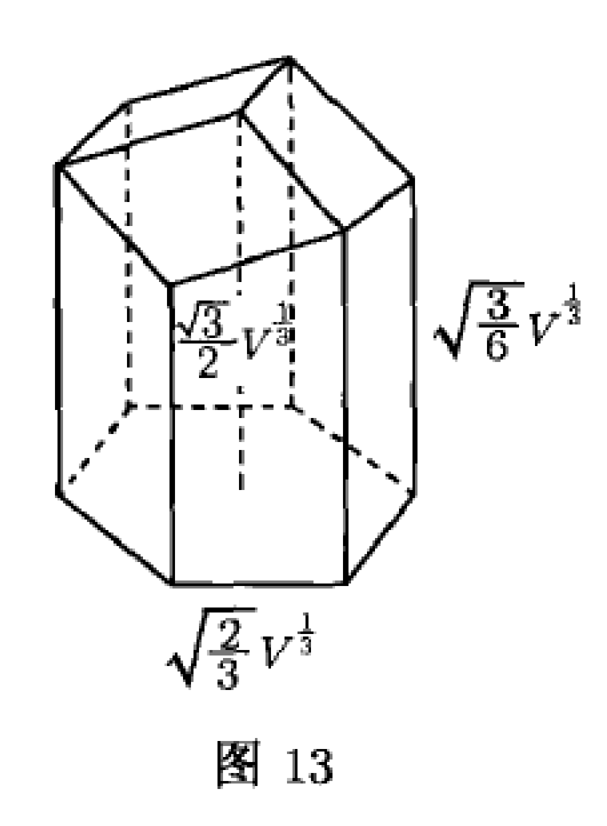



图 13

**证** 由第四节已知尖顶六棱柱的体积 $V$ 和表面积 $S$ 各为

$$V=\frac{3\sqrt{3}}{2}a^2b,$$

$$S=6a\left(b+\frac{a}{\sqrt{8}}\right),$$

即

$$S=\frac{4V}{\sqrt{3}a}+\frac{6}{\sqrt{8}}a^2=\frac{2V}{\sqrt{3}a}+\frac{2V}{\sqrt{3}a}+\frac{6}{\sqrt{8}}a^2.$$

由公式 (1) 得出

$$S\geqslant 3\left(\frac{2V}{\sqrt{3}a}\times\frac{2V}{\sqrt{3}a}\times\frac{6}{\sqrt{8}}a^2\right)^{\frac{1}{3}}=3\sqrt{2}V^{\frac{2}{3}};$$

而且仅当

$$\frac{2V}{\sqrt{3}a}=\frac{6}{\sqrt{8}}a^2,$$

也就是

$$a=\sqrt{\frac{2}{3}}V^{\frac{1}{3}},\ b=\frac{1}{\sqrt{3}}V^{\frac{1}{3}}$$

时 $S$ 取最小值. 但必须检验这是否适合于条件

$$b \geqslant \frac{1}{\sqrt{8}}a;$$

如果不适合, 可能出现六节所指出的情况, 而这个数值是不能达到的.

这尖顶六棱柱的高度是 $b+\frac{1}{\sqrt{8}}a=\frac{1}{2}\sqrt{3}V^{\frac{1}{3}}$, 它的棱长高的是 $b=\frac{1}{\sqrt{3}}V^{\frac{1}{3}}$, 低的是 $b-\frac{1}{\sqrt{8}}a=\frac{1}{6}\sqrt{3}V^{\frac{1}{3}}$.

**定理 2**　体积为 $V$ 的屋脊四方柱的表面积 (不算底面) 的最小值是 $V^{\frac{2}{3}}$, 而且仅当正方形边长是 $\frac{2^{2/3}}{3^{1/6}}V^{\frac{1}{3}}$ 及檐高 $\frac{1}{2^{1/3}3^{2/3}}V^{\frac{1}{3}}$ 的情况下取这最小值.

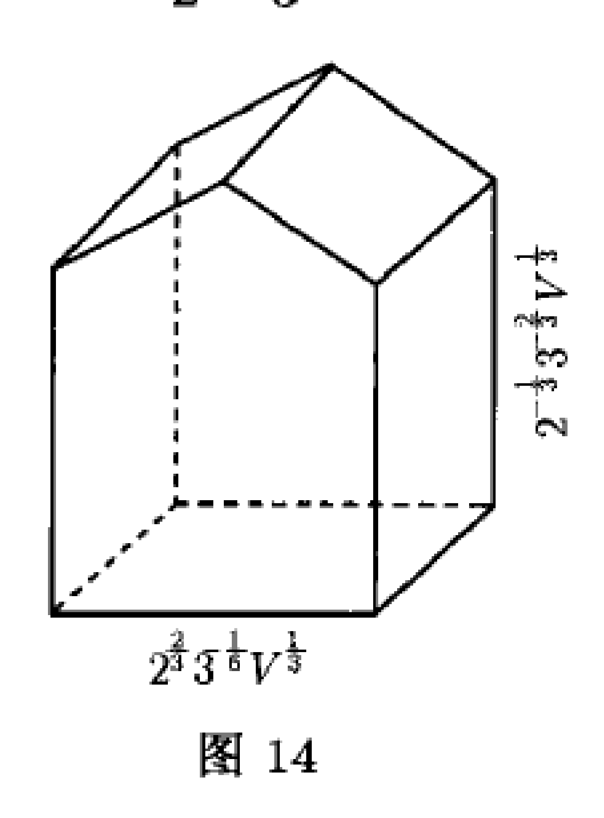



图 14

**证**　由七节已知屋脊四方柱的体积 $V$ 和表面积 $S$ 各等于

$$V=a^2b,$$

$$S=4a\left(b+\frac{\sqrt{3}}{8}a\right),$$

即

$$S=\frac{4V}{a}+\frac{\sqrt{3}}{2}a^2=\frac{2V}{a}+\frac{2V}{a}+\frac{\sqrt{3}}{2}a^2.$$

由公式 (1) 得出

$$S \geqslant 3\left(\frac{2V}{a}\times\frac{2V}{a}\times\frac{\sqrt{3}}{2}a^2\right)^{\frac{1}{3}}=2^{\frac{1}{3}}3^{\frac{7}{6}}V^{\frac{2}{3}};$$

而且仅当

$$\frac{2V}{a}=\frac{\sqrt{3}}{2}a^2,$$

也就是

$$a=\frac{2^{2/3}}{3^{1/6}}V^{\frac{1}{3}},\ b=\frac{3^{1/3}}{2^{4/3}}V^{\frac{1}{3}}$$

时 $S$ 取最小值. 易见 $b>\frac{a}{4\sqrt{3}}$. 因此檐高等于

$$b-\frac{1}{4\sqrt{3}}a=\left(\frac{3^{1/3}}{2^{4/3}}-\frac{1}{2^{4/3}3^{2/3}}\right)V^{\frac{1}{3}}=\frac{1}{2^{1/3}3^{2/3}}V^{\frac{1}{3}};$$

脊高为 $b+\frac{1}{4\sqrt{3}}a=\frac{2^{2/3}}{3^{2/3}}V^{\frac{1}{3}}$. 截面如图 15, 即六角形的一半.

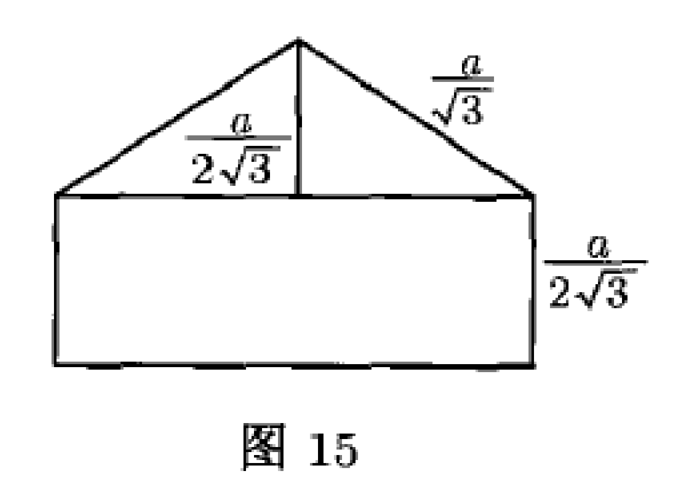



图 15

**结论** 由于

$$3\sqrt{2}<3^{\frac{7}{6}}2^{\frac{1}{3}},$$

所以在保证同样容量的条件下, 尖顶六棱柱比屋脊四方柱用材要少.

用同样方法不难证明, 体积为 $V$ 的尖顶四方柱的表面积 (不算底面) 的最小值是 $3\sqrt{2}V^{\frac{2}{3}}$, 而且仅当正方形边长为 $\sqrt{2}V^{\frac{1}{3}}$ 及檐高为“0”的情况下取这最小值. 说得更清楚些, 这是个以 $\sqrt{2}V^{\frac{1}{3}}$ 为底边长, 以 $\left(2\times\frac{1}{\sqrt{8}}\times\sqrt{2}V^{\frac{1}{3}}=\right)V^{\frac{1}{3}}$ 为高的尖顶形, 或即将菱形十二面体拦腰一截, 所得之半. 因此在同样容量下, 这种容器和尖顶六棱柱用材相同.

虽然如此, 但实际测量一下, 蜂房的大小与定理 1 中所给出的比例并不相合. 经过实测, $a\doteq 0.35$ 厘米, 深 $b+\frac{1}{\sqrt{8}}a\doteq 0.70$ 厘米, 而按定理 1, $b+\frac{1}{\sqrt{8}}a=\frac{1}{\sqrt{2}}a+\frac{1}{\sqrt{8}}a=\frac{3}{\sqrt{8}}\times 0.35\doteq 0.38$ 厘米.

正是: 往事几百年, 祖述前贤, 瑕疵讹谬犹盈篇, 蜂房秘奥未全揭, 待我向前!

让我们再看看, 添上一扇以底面作的“门”, 问哪种形状最好?

先看屋脊四方柱, 它是在体积为

$$V=a^2b$$

的情况下求表面积 (包括“门”在内)

$$S=4a\left(b+\frac{\sqrt{3}}{8}a\right)+a^2$$

的最小值. 由

$$S=\frac{4V}{a}+\left(\frac{\sqrt{3}}{2}+1\right)a^2\geqslant 3\left[\frac{2V}{a}\times\frac{2V}{a}\left(\frac{\sqrt{3}}{2}+1\right)a^2\right]^{\frac{1}{3}}$$

$$=3[2(\sqrt{3}+2)]^{\frac{1}{3}}V^{\frac{2}{3}},$$

并且仅当

$$\frac{2V}{a}=\left(\frac{\sqrt{3}}{2}+1\right)a^2$$

时取等号, 即

$$a=\left[\frac{4}{\sqrt{3}+2}\right]^{\frac{1}{3}}V^{\frac{1}{3}}=[4(2-\sqrt{3})]^{\frac{1}{3}}V^{\frac{1}{3}}$$

时取等号. 其时

$$b=\frac{1}{[4(2-\sqrt{3})]^{2/3}}V^{\frac{1}{3}}=\left(\frac{2+\sqrt{3}}{4}\right)^{\frac{2}{3}}V^{\frac{1}{3}}.$$

再看尖顶六棱柱. 它是在体积

$$V=\frac{3\sqrt{3}}{2}a^2b$$

的情况下求表面积 (包括"门"在内)

$$S=6a\left(b+\frac{a}{\sqrt{8}}\right)+\frac{3\sqrt{3}}{2}a^2$$

的最小值. 由

$$\begin{aligned}S=&2\times\frac{2}{\sqrt{3}}\times\frac{V}{a}+\left(\frac{3}{\sqrt{2}}+\frac{3\sqrt{3}}{2}\right)a^2\\ \geqslant&3\left[\frac{2}{\sqrt{3}}\frac{V}{a}\times\frac{2}{\sqrt{3}}\frac{V}{a}\left(\frac{3}{\sqrt{2}}+\frac{3\sqrt{3}}{2}\right)a^2\right]^{\frac{1}{3}}\\ =&3[2(\sqrt{2}+\sqrt{3})]^{\frac{1}{3}}V^{\frac{2}{3}},\end{aligned}$$

且仅当

$$\frac{2V}{\sqrt{3}a}=\left(\frac{3}{\sqrt{2}}+\frac{3\sqrt{3}}{2}\right)a^2,$$

即

$$\begin{aligned}a=&\frac{4^{1/3}}{\sqrt{3}}(\sqrt{3}-\sqrt{2})^{\frac{1}{3}}V^{\frac{1}{3}},\\ b=&\frac{2}{3\sqrt{3}}\times\frac{3}{(\sqrt{3}-\sqrt{2})^{2/3}}\times\frac{1}{2^{4/3}}V^{\frac{1}{3}}\\ =&\frac{1}{2^{1/3}\times\sqrt{3}(\sqrt{3}-\sqrt{2})^{2/3}}V^{\frac{1}{3}}\end{aligned}$$

时取等号.

两下相比, 由于

$$3[2(\sqrt{3}+2)]^{\frac{1}{3}} > 3[2(\sqrt{3}+\sqrt{2})]^{\frac{1}{3}}.$$

所以还是尖顶六棱柱来得好.

对于尖顶四方柱而言, 可以算出表面积 (包括“门”在内) 的最小值为 $3[2(2+\sqrt{2})]^{\frac{1}{3}}V^{\frac{2}{3}}$, 也没有尖顶六棱柱来得好.

对于尖顶六棱,

$$\frac{a}{b}=\frac{2}{\sqrt{3}+\sqrt{2}}\doteq 0.64,$$

与实测所得的 $\frac{a}{b}\doteq\frac{0.35}{0.58}\doteq 0.6$ 相比相当接近. 有没有道理?

## 九　正　　题

由上可知, 客观情况并不单纯是一个“体积给定, 求用材最小”的数学问题, 那样的提法是不妥当的. 现在让我们来重提看看.

把蜜蜂的体态入算. 从考虑它的身长、腰围入手, 怎样情况用材最省?

首先, 那尖顶六棱柱所能容纳的“腰围”等于 $\sqrt{3}a$(图 16), 长度是 $b+\frac{1}{\sqrt{8}}a$. 另一方面屋脊四方柱的“腰围”等于 $a_1$, 长度等于 $b_1+\frac{1}{4\sqrt{3}}a_1$. 让我们在粗长各相等, 即在

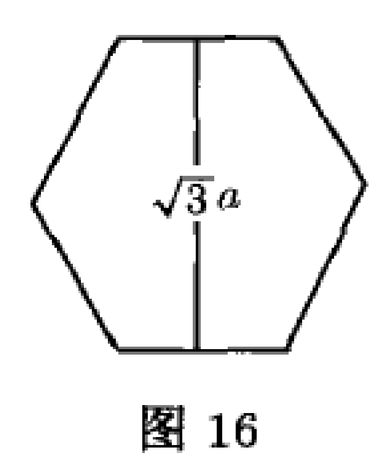



图 16

$$\sqrt{3}a=a_1,$$
$$b+\frac{1}{\sqrt{8}}a=b_1+\frac{1}{4\sqrt{3}}a_1$$

的条件下考虑问题. 由于

$$\begin{aligned}S_1=&4a_1\left(b_1+\frac{\sqrt{3}}{8}a_1\right)\\=&4a_1\left[\left(b_1+\frac{1}{4\sqrt{3}}a_1\right)+\left(\frac{\sqrt{3}}{8}-\frac{1}{4\sqrt{3}}\right)a_1\right]\\>&4a_1\left(b_1+\frac{1}{4\sqrt{3}}a_1\right)=4\sqrt{3}a\left(b+\frac{1}{\sqrt{8}}a\right)\end{aligned}$$

$$>6a\left(b+\frac{1}{\sqrt{8}}a\right)=S.$$

即在同长同粗的情况下, 尖顶六棱柱比屋脊四方柱省料些.

这建议了以下的猜测:

量体裁衣, 形状为尖顶六棱柱的蜂房, 是最省材料的结构, 它比屋脊四方柱还要节省材料.

再看带"门"的情况. 仍然

$$\sqrt{3}a=a_1,\ b+\frac{1}{\sqrt{8}}a=b_1+\frac{1}{4\sqrt{3}}a_1.$$

但需要比较

$$S_1=4a_1\left(b_1+\frac{\sqrt{3}}{8}a_1\right)+a_1^2$$

与 $S=6a\left(b+\frac{1}{\sqrt{8}}a\right)+\frac{3\sqrt{3}}{2}a^2=6ab+\left(\frac{3}{\sqrt{2}}+\frac{3\sqrt{3}}{2}\right)a^2$ 谁大. 以 $a_1=\sqrt{3}a$, $b_1=b+\frac{1}{\sqrt{8}}a-\frac{1}{4}a$ 代入 $S_1$, 得

$$\begin{aligned}S_1=&4\sqrt{3}a\left(b+\frac{1}{\sqrt{8}}a-\frac{1}{4}a+\frac{3}{8}a\right)+3a^2\\=&4\sqrt{3}ab+\left(\sqrt{6}+\frac{\sqrt{3}}{2}+3\right)a^2\\>&6ab+\left(\frac{\sqrt{3}}{2}+\frac{3\sqrt{3}}{2}\right)a^2=S.\end{aligned}$$

也就是说, 带上门, 还是蜂窝来得好.

这说明了生物本身与环境的关系的统一性.

**附记** 读者不难证明, 如果我们考虑 $x,y$ 轴刻度不一致的正六边形 (图 17), 考虑由此所作出的六棱柱和尖顶六棱柱, 我们不难证明, 在体积给定的条件下, 仍然以第八节中所得出的图形表面积最小.

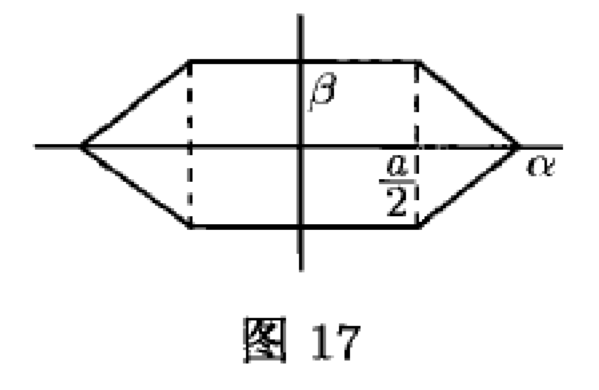



图 17

学过微积分的读者可以看出, 我实在是在"分散难点". 确切地说, 这是一个四个变数求条件极值的问题. 四个变数是指 $x,y,z$ 轴各增加若干倍, 并在某点切下来; 条件是等体积.

## 十 设 问

从以上所谈的一些情况看来, 我们只不过从六棱柱 (或四方柱) 出发, 按一定的切拼方法做了些研究而已. 实质上, 这样的看法未入事物之本质. 为什么仅从六棱柱出发, 而不能从三角柱、四方柱或其他柱形出发, 甚至于为什么要从柱形出发? 更不要说切拼之法也是千变万化了! 甚至于为什么要从切拼得来! 越想问题越多, 思路越宽.

把两个蜂房门对门地联接起来, 得出以下两种可能的图形 (图 18、19).

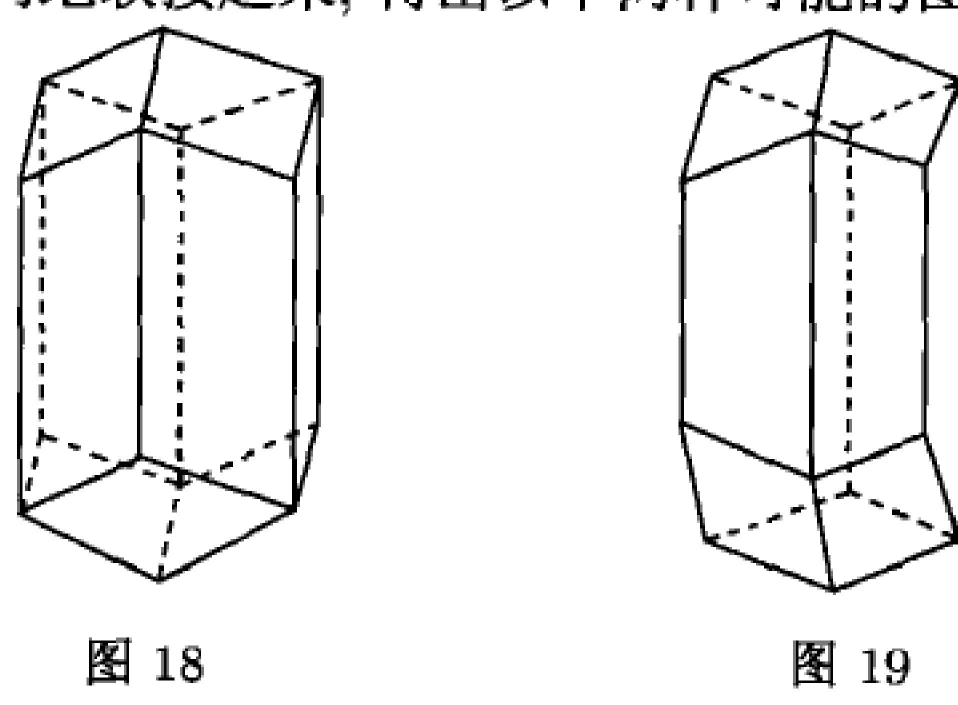


图 18　　图 19

这两图形都有以下的性质: 用这种形式的砖可以填满整个空间. 有这样性质的砖, 就是结晶体. 图 19 所表达的其实就是透视石的晶体.

两个屋脊四棱柱口对口地接在一起, 两个尖顶四棱柱口对口地接在一起, 各得黄赤沸石 (图 20) 与锆英石 (图 21) 的晶体图形. 特别, 两个尖顶形口对口地接在一起得一菱形十二面体, 也就是石榴子石晶体的图形 (图 22).

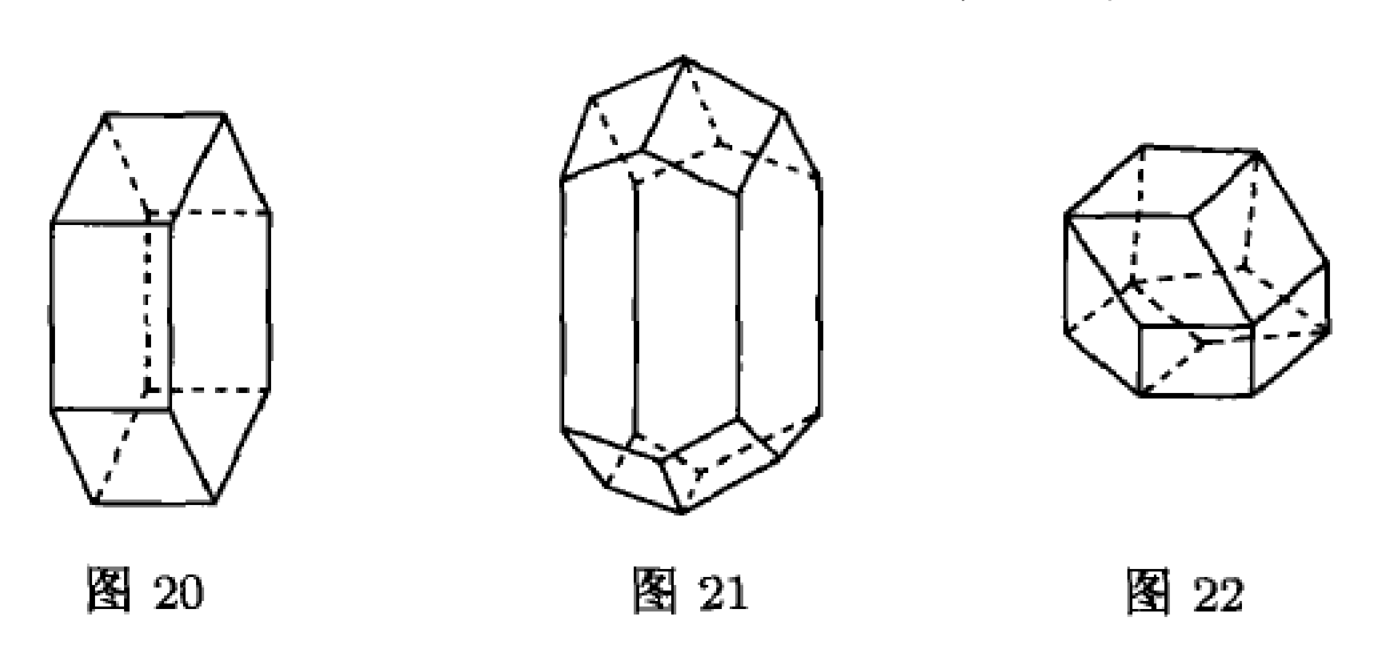


图 20　　图 21　　图 22

因而归纳出以下的基本问题:

**问题 1**　怎样的体可以作为晶体? 也就是说, 用同样的体可以无穷无尽地, 无空无隙地填满整个空间.

这是有名的晶体问题. 经过费德洛夫的研究知道, 晶体可分为 230 类.

**问题 2**　给定体积, 哪一类晶体的表面积最小?

**问题 3**　给一个一定的体形, 求出能包有这体形的表面积最小的晶体. 例如,

图 23 给了一个橄榄或一个陀螺, 求包这橄榄或陀螺的表面积最小的晶体. 把那晶体拦腰切为两段, 那可能是蜂房的最佳结构了.

为了补充些感性知识, 我们再讲些例子.

柱体填满空间的问题等价于怎样的样板可以填满平面的问题. 以任何一个三角形为样板都可以填满平面 (图 24). 任何四边形也可以作为样板用来填满平面 (图 25). 一个正六边形 (或六个角都是 120° 的图形), 也可以用来作为样板填满平面 (图 26).

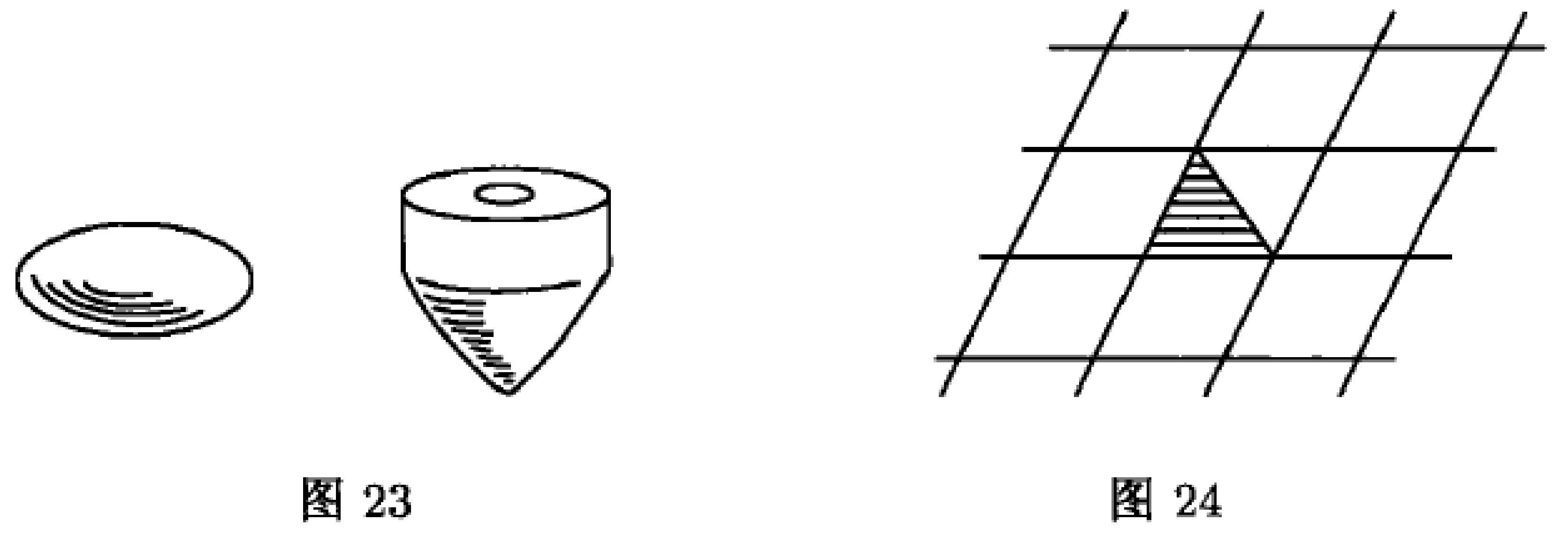


图 23　　图 24

在这三个图形中可以看出什么公共性质? 例如图 24 的各边中点形成怎样的网格; 在图 26 中联上六边形的三条"对角线"得出怎样的图形?

为什么要求的只是填上整个空间或平面, 而不是一个球或一个圆柱?

现在让我们以球为例来作一些探讨.

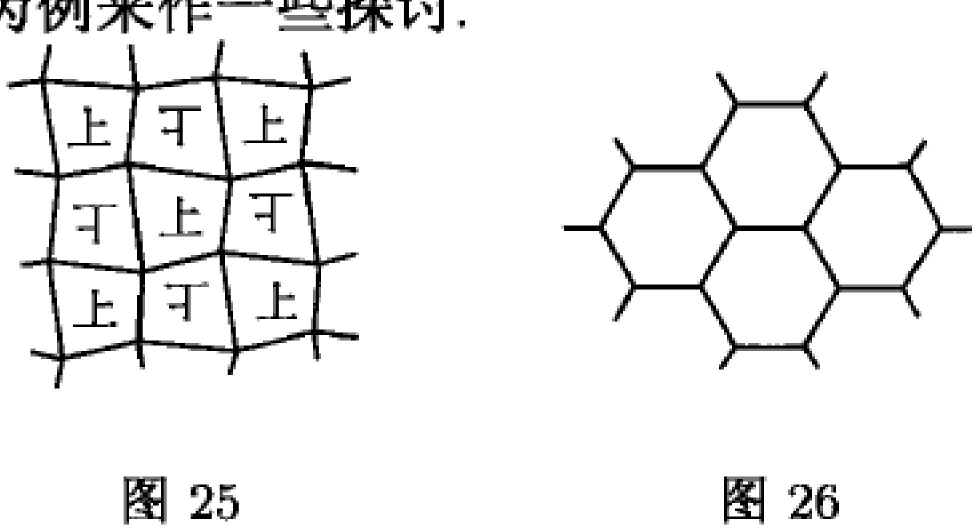


图 25　　图 26

**例 1**　把球心为二十四等分. 以球分为原点, 引三条坐标轴, 将球分为八等分 (八个卦限, 每个卦限一份) 如图 27; 再从每片球面的中心向三顶点如图划分, 共得 24 份.

**例 2**　把球分为六十等分. 从球内接正二十面体 (图 28) 出发, 向球上投影得 20 个三角形; 再把每一个三角形依中心到三角的联线分为三等分, 因而共有 60 份.

**例 3**　把柱形直切成四等分; 再切成片, 每一个成一间房 (图 29). 另一方法, 作柱上开口像六角形的图形拼上 (图 30).

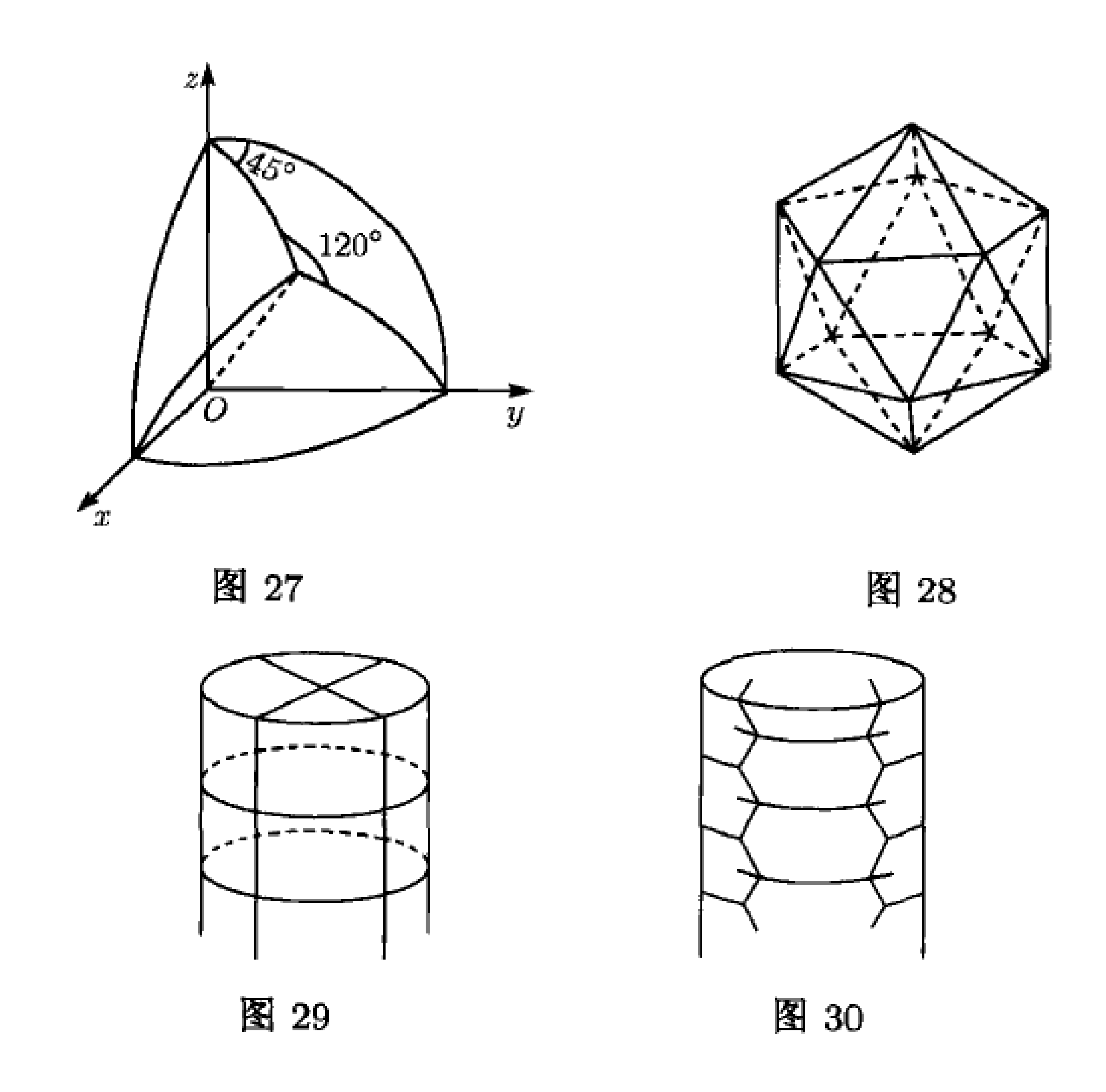



图 27

图 28

图 29

图 30

(感谢许杰教授, 他给我看到了古生物笔石的模型, 启发出这一图象.)

由蜂房启发出来的问题, 所联想到的问题何止于此! 浮想联翩, 由此及彼, 花样真不少呢! 据说飞机的蜂窝结构也是由此而启发出来的, 但可以根据不同的要求, 而得出各种各样的求极值问题来 (例如, 使结构的强度达到最高的问题). 就数学来说, 由此可以想到的问题真也不少. 本文是为高中水平的读者写的, 因而不能不适可而止了. 但是为了让同学们在进入大学之后还可以咀嚼一番, 回味一番, 我在此后再添讲几节. 一来看看怎样"浮想"; 二来给同学提供一个例子: 怎样从一个问题而复习我们所学到的东西, 这样复习就使有些学过的内容自然而然地串联起来了.

## 十一　代　　数

在阅读十二、十三节等内容以前, 我们先来一段插话. 这段插话可以不看. 也许看了一下会觉得有些联系不上, 但将来回顾一下, 读者会有深长体会的!

经过旋转, 平移, 透视石 (两个蜂房对合所成的图形) 的表面积和体积不变. 在第九节中曾经提出过把 $x, y, z$ 轴各增长若干倍而看一个透视石体积及表面积的变化的情况. 如果体积不变, 怎样的倍数才能使表面积取最小值? 这实质上是求: 在群

$$x' = \alpha_1 x + \beta_1 y + \gamma_1 z + \delta_1, y' = \alpha_2 x + \beta_2 y + \gamma_2 z + \delta_2,$$

$$z' = \alpha_3 x + \beta_3 y + \gamma_3 z + \delta_3,$$

$$\left(\begin{vmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 \\ \alpha_3 & \beta_3 & \gamma_3 \end{vmatrix} = 1\right)$$

下, 等价于一个透视石的诸图形中, 哪一个表面积最小.

看来, 对平行六面体的讨论可能容易些. 用无数个同样的平行六面体可以填满空间. 如果六面体的体积给了, 怎样的形状表面积最小 (或棱的总长最短, 或棱的长度的乘积最小)?

讲到这儿, 暂且摆下, 慢慢咀嚼, 慢慢体会这一段话与以下所讲的东西的关系. 我们先看一批代数不等式.

**例 1** 求证

$$2\sqrt{|ad-bc|} \leqslant \sqrt{a^2+b^2}+\sqrt{c^2+d^2}, \tag{1}$$

而且仅当 $\frac{a}{b}=\frac{d}{c}$ 及 $|b|=|c|$ 或 $|a|=|b|$ 时取等号.

**证** 由

$$\begin{aligned}(a^2+b^2)(c^2+d^2) &= (ad-bc)^2+(ac+bd)^2 \\ &\geqslant (ad-bc)^2,\end{aligned}$$

因此

$$|ad-bc| \leqslant \sqrt{(a^2+b^2)(c^2+d^2)}, \tag{2}$$

$$\begin{aligned}\sqrt{|ad-bc|} &\leqslant \sqrt{\sqrt{a^2+b^2}\sqrt{c^2+d^2}} \\ &\leqslant \frac{1}{2}(\sqrt{a^2+b^2}+\sqrt{c^2+d^2}),\end{aligned}$$

即得所证.

**例 2** 求证

$$\begin{aligned}6\sqrt{|ab-bc|} \leqslant\ & 2\sqrt{a^2+c^2} \\ &+\sqrt{a^2+c^2+3(b^2+d^2)-2\sqrt{3}(ab+cd)} \\ &+\sqrt{a^2+c^2+3(b^2+d^2)+2\sqrt{3}(ab+cd)}.\end{aligned} \tag{3}$$

读者试自己证明此式, 并且试证以下两不等式. 最好等证毕后再看十三节.

**例 3** 求证

$$\begin{aligned}16|ad-bc|^3 \leqslant & (a^2+c^2)\{[a^2+c^2+3(b^2+d^2)]^2 \\ &-12(ab+cd)^2\}.\end{aligned} \tag{4}$$

更一般些, 有

**例 4** 当 $n \geqslant 1$ 时,

$$|ad-bc|^n \leqslant \prod_{l=1}^{n}\left[(a^2+c^2)\sin^2\frac{\pi(2l-1)}{n} -2(ab+cd)\sin\frac{\pi(2l-1)}{n}\cos\frac{\pi(2l-1)}{n} +(b^2+d^2)\cos^2\frac{\pi(2l-1)}{n}\right], \tag{5}$$

或

$$|ad-bc|^{\frac{1}{2}} \leqslant \frac{1}{n}\sum_{l=1}^{n}\left[(a^2+c^2)\sin^2\frac{\pi(2l-1)}{n} -2(ab+cd)\sin\frac{\pi(2l-1)}{n}\cos\frac{\pi(2l-1)}{n} +(b^2+d^2)\cos^2\frac{\pi(2l-1)}{n}\right]^{\frac{1}{2}}, \tag{6}$$

(5) 式比 (6) 式难些, 我们将在十四节中给以证明.

## 十二 几 何

看看上节 (1) 式及 (2) 式的几何意义如何? 在平面上作三点 $O(0,0), A(a,b)$ 及 $B(c,d)$. 以 $OA, OB$ 为边的平行四边形的面积等于 $|ad-bc|, OA, OB$ 的长度各为 $\sqrt{a^2+b^2}, \sqrt{c^2+d^2}$. 所以上节不等式 (2) 的意义是*平行四边形的面积小于或等于两邻边的乘积*.

而不等式 (1) 的意义是: 平行四边形面积的平方根小于或等于其周长的四分之一, 即四边长的平均值, 并且仅当正方形时取等号; 或者说 (1) 的意义也就是*周长一定的平行四边形中, 以正方形的面积为最大*.

再看不等式 (3), 从代数的角度来看有些茫然, 有些突然. 但从几何来看却是*“周长一定的六边形中, 以正六角形的面积为最大”*的这一性质的特例. 不等式 (6) 也可以作如是观.

这些结果的几何意义是明显的. 但如果抛开几何, 就式论式, 从代数角度来看就有些意外了. 实质岂其然哉, 谬以千里! 那不过是几何的性质用代数的语言说出而已.

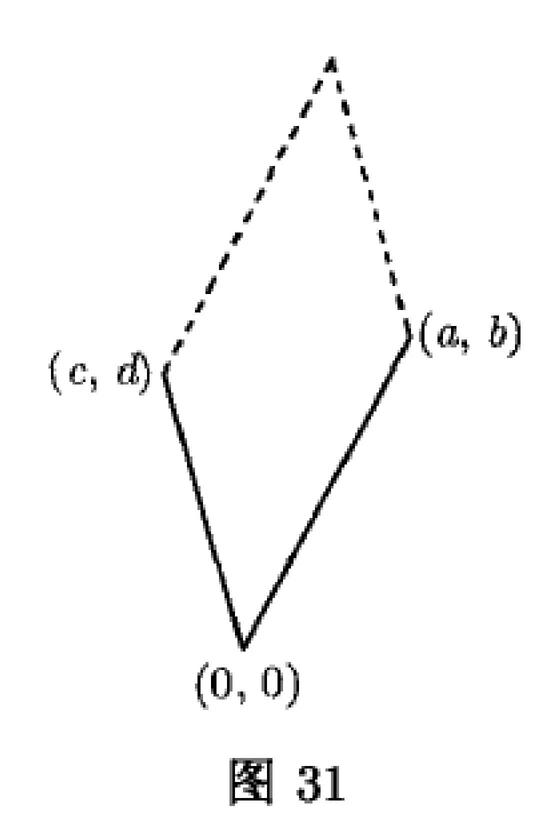



图 31

这是“几何”启发出“代数”，但代数的考虑又大大地丰富了几何. 由于几何平均小于算术平均，因此，由 (6) 式可以追问更精密的 (5) 式对不对. 要从几何角度直接看出 (5) 式来是不太容易的.

不要说不等式 (2) 简单，推广到 $n$ 维空间就有

$$\begin{vmatrix} a_{11}, & \cdots, & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1}, & \cdots, & a_{nn} \end{vmatrix} \leqslant \sum_{i=1}^{n} a_{1i}^2 \sum_{i=1}^{n} a_{2i}^2 \cdots \sum_{i=1}^{n} a_{ni}^2.$$

这是有名的 Hadamard 不等式，但其几何直观已尽乎此点，对有丰富几何直观的人来说此式之发明并非出人意料了. 正是：

**数与形，本是相倚依，焉能分作两边飞. 数缺形时少直觉，形少数时难入微. 数形结合百般好，隔裂分家万事非. 切莫忘，几何代数统一体，永远联系，切莫分离!**

## 十三　推　　广

我们现在来证明十一节中公式 (5). 在证明之前，我们换一下符号. 命

$$A = a^2 + c^2, B = -ab - cd, C = b^2 + d^2,$$

因此

$$(ad - bc)^2 = (a^2 + c^2)(b^2 + d^2) - (ab + cd)^2 = AC - B^2.$$

十一节的公式 (5) 一变而为求证：如果 $A > 0, AC - B^2 > 0$，则

$$(AC - B^2)^{\frac{n}{2}} \leqslant \sum_{l=1}^{n} \left[ A \sin^2 \frac{\pi(2l-1)}{n} \right.$$
$$\left. + 2B \sin \frac{\pi(2l-1)}{n} \cos \frac{\pi(2l-1)}{n} \right.$$

$$+C\cos^2\frac{\pi(2l-1)}{n}\Bigg]. \tag{1}$$

这式子的右边用 $P$ 表之, 用倍角公式得

$$\begin{aligned}P=&\prod_{l=1}^{n}\left[\frac{1}{2}(A+C)+B\sin\frac{2\pi(2l-1)}{n}\right.\\&\left.+\frac{1}{2}(C-A)\cos\frac{2\pi(2l-1)}{n}\right]\\=&\prod_{l=1}^{n}\left[p-q\cos\left(\frac{2\pi(2l-1)}{n}+\eta\right)\right],\end{aligned}$$

这儿

$$p=\frac{1}{2}(A+C),\ q=\sqrt{B^2+\frac{1}{4}(C-A)^2}, \tag{2}$$

及

$$\sin\eta=\frac{1}{2}(C-A)/q.$$

考虑

$$\begin{aligned}&\left[u-v\exp\left(\frac{2\pi(2l-1)i}{n}+\eta i\right)\right]\left[u-v\exp\left(-\frac{2\pi(2l-1)i}{n}-\eta i\right)\right]\\&=u^2+v^2-2uv\cos\left(\frac{2\pi(2l-1)}{n}+\eta\right),\end{aligned}$$

这儿 $e^x$ 写成为 $\exp x$, 如果取得 $u,v$ 使

$$u^2+v^2=p,\ 2uv=q, \tag{3}$$

则 $P$ 可以写成为 $Q\bar{Q}$, 其中

$$Q=\prod_{l=1}^{n}(u-ve^{\eta i}e^{2\pi i(2l-1)/n}).$$

当 $n$ 是奇数时,

$$Q=\prod_{m=1}^{n}(u-ve^{\eta i}e^{2\pi im/n})=u^n-v^ne^{ni\eta},$$

即得

$$\begin{aligned}P=&(u^n-v^ne^{ni\eta})(u^n-v^ne^{-ni\eta})\\=&u^{2n}+v^{2n}-2u^nv^n\cos n\eta;\end{aligned} \tag{4}$$

当 $n$ 是偶数时,

$$
\begin{aligned}
Q=&\sum_{l=1}^{\frac{n}{2}}(u-ve^{\eta i}e^{-2\pi in}e^{2nl/\frac{n}{2}})^2\\
=&[u^{\frac{n}{2}}-(ve^{\eta i}e^{-2\pi in})^{\frac{n}{2}}]^2=(u^{\frac{n}{2}}+v^{\frac{n}{2}}e^{n\eta i/2})^2.
\end{aligned}
$$

即得

$$
P=\left(u^n+v^n+2u^{\frac{n}{2}}v^{\frac{n}{2}}\cos\frac{n\eta}{2}\right)^2. \tag{5}
$$

我们现在需要以下的简单引理.

**引理**　当 $u>v\geqslant 0$ 及 $m\geqslant 2$ 时,

$$
(u^m-v^m)^2\geqslant(u^2-v^2)^m. \tag{6}
$$

这引理等价于, 当 $1>x>0$ 时,

$$
(1-x^m)^2\geqslant(1-x^2)^m.
$$

**证**　由于 $x^m>x^{m+1}$, 所以, 若原式成立, 应有

$$
(1-x^{m+1})^2>(1-x^m)^2>(1-x^2)^m(1-x^2)=(1-x^2)^{m+1}.
$$

原式在 $m=2$ 时显然成立. 因此, 由数学归纳法可以证明 (6) 式

把这引理用到 (4) 式, 当 $n$ 是奇数时,

$$
P\geqslant u^{2n}+v^{2n}-2u^nv^n=(u^n-v^n)^2\geqslant(u^2-v^2)^n.
$$

由 (3) 及 (2) 可知

$$
\begin{aligned}
(u^2-v^2)^2=p^2-q^2=&\frac{1}{4}(A+C)^2-B^2-\frac{1}{4}(A-C)^2\\
=&AC-B^2,
\end{aligned}
$$

即得 (1) 式.

当 $n$ 是偶数时, 由 (5) 式

$$
\begin{aligned}
P\geqslant&(u^n+v^n-2u^{\frac{n}{2}}v^{\frac{n}{2}})^2=(u^{\frac{n}{2}}-v^{\frac{n}{2}})^4\\
\geqslant&(u^2-v^2)^n=(AC-B^2)^{\frac{n}{2}},
\end{aligned}
$$

也得 (1) 式.

**附记 1** 当 $m > 2$ 时, 引理中的不等式, 当且仅当 $v = 0$ 时取等号, 而 $v = 0$ 等价于

$$q = B^2 + \frac{1}{4}(C - A)^2 = 0,$$

即当且仅当 $B = 0, A = C$ 时, (1) 式取等号. 但须注意 $n = 4$ 的情况必须除外 (因为 $m = 2$ 了), 这时取等号的情形应当是 $\cos 2\eta = -1$, 即 $\eta=90°$ , 及 $A = C$.

回到原来的问题, 当 $n \neq 4$ 时,

$$a^2 + c^2 = b^2 + d^2,\ ab + cd = 0, \tag{7}$$

即十一节 (6) 式成立, 而且仅当 (7) 式成立时取等号.

**附记 2** 当 $n = 4$ 时, 不等式十一节 (5) 是

$$(ad - bc)^2 \leqslant \frac{1}{4}[(a+b)^2 + (c+d)^2][(a-b)^2 + (c-d)^2],$$

当且仅当 $a^2 + c^2 = b^2 + d^2$ 时取等号. 换符号

$$\alpha = a + b,\ \beta = a - b,\ \gamma = c + d,\ \delta = c - d,$$

则得

$$(\alpha\delta - \beta\gamma)^2 \leqslant (\alpha^2 + \gamma^2)(\beta^2 + \delta^2),$$

当且仅当 $a^2 + c^2 - b^2 - d^2 = \alpha\beta + \gamma\delta = 0$ 时取等号. 这就是 Hadamard 不等式.

这样的推广实际上可以说“退了一步”的推广. 我们原来所讨论的问题是三维的, 而我们退成二维, 然后看看二维中有哪些推广的可能性. 这是一般的研究方法: 先足够地退到我们所最容易看清楚问题的地方, 认透了钻深了, 然后再上去.

我们再看另外一个例子, 在这例子中我们希望把三角形, 四面体 $\cdots\cdots$ 推广到 $n$ 维空间的 $(n+1)$ 面体的问题.

在 $n$ 维空间取 $n+1$ 点, 这 $n+1$ 点中的任意 $n$ 点决定一平面, 共有 $n+1$ 个面, 这些面包有的体的容积是 $V$. 过 $n+1$ 点中的任意两点可以作一线段, 共有 $\frac{1}{2}n(n+1)$ 条线段. 我们的问题是 $V$ 给定了, 求这 $\frac{1}{2}n(n+1)$ 条线段乘积的最小值.

读者不要小看这问题, 自己试试“四面体”便知其分量了.

## 十四 极 限

把十一节的 (6) 式推到 $n \to \infty$ 的情形. 由积分的定义

$$|ad-bc|^{\frac{1}{2}} \leqslant \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n} \left[(a^2 + c^2)\sin^2 \frac{\pi(2l-1)}{n}\right.$$

$$-2(ab+cd)\sin\frac{\pi(2l-1)}{n}\cos\frac{\pi(2l-1)}{n}$$
$$\left.+(b^2+d^2)\cos^2\frac{\pi(2l-1)}{n}\right]^{\frac{1}{2}}$$
$$=\frac{1}{2\pi}\int_0^{2\pi}[(a\ \sin\ \theta-b\ \sin\ \theta)^2+(c\ \sin\ \theta-d\ \cos\ \theta)^2]^{\frac{1}{2}}d\theta.$$

换符号 $a^2+c^2=A,\ B=-ab-cd,\ C=b^2+d^2$, 则得

$$(AC-B^2)^{\frac{1}{4}}\leqslant\frac{1}{2\pi}\int_0^{2\pi}(A\ \sin^2\ \theta+2B\ \sin\ \theta\ \cos\ \theta+C\ \cos^2\ \theta)^{\frac{1}{2}}d\theta. \tag{1}$$

从十一节 (5) 可以得出更精密的不等式

$$\frac{1}{2}\log(AC-B^2)\leqslant\frac{1}{2\pi}\int_0^{2\pi}\log(A\ \sin^2\ \theta+2B\ \sin\ \theta\ \cos\ \theta+C\ \cos^2\ \theta)d\theta. \tag{2}$$

我们能不能直接证明这些不等式? 能! 读过微积分的读者自己想想看.

## 十五　抽　　象

在平面上给出一块样板 $N$(图 32), 变换

$$(T)\begin{cases}\xi=ax+by+p,\\ \eta=cx+dy+q.\end{cases}\quad ad-bc\neq 0.$$

把 $N$ 变成为 $(\xi,\eta)$ 平面上的 $N(T)$. $N(T)$ 的面积及周界长度各命之为 $A(T)$ 与 $L(T)$, 求

$$\frac{(A(T))^{\frac{1}{2}}}{L(T)}$$

的最大值.

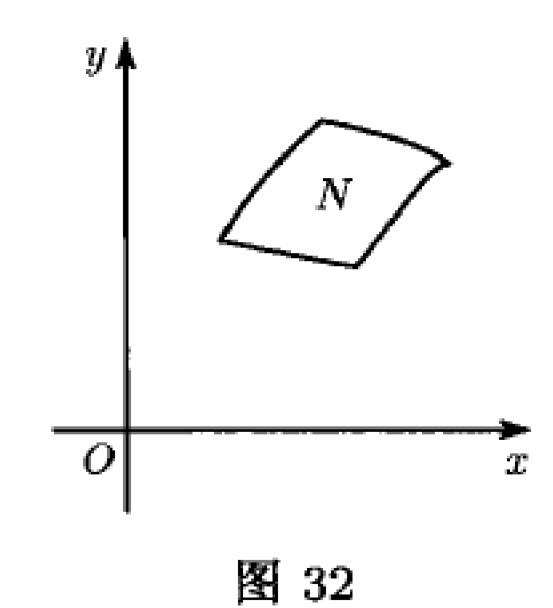



图 32

取单位圆内接正 $n$ 角形为样板, 它的周长和面积各为 $2n\sin\frac{\pi}{n}$ 与 $\frac{n}{2}\sin\frac{2\pi}{n}$. 不妨假定正 $n$ 边形的顶点就是

$$\left(\cos\frac{2\pi l}{n},\ \sin\frac{2\pi l}{l}\right),\quad l=0,1,2,\cdots,n-1.$$

经变换 $(T)$ 后, 这 $n$ 点各变为

$$\begin{cases}\xi_l=a\cos\dfrac{2\pi l}{n}+b\sin\dfrac{2\pi l}{n}+p,\\ \eta_l=c\cos\dfrac{2\pi l}{n}+d\sin\dfrac{2\pi l}{n}+q.\end{cases}$$

第 $l$ 边的长度是

$$\sqrt{(\xi_l-\xi_{l-1})^2+(\eta_l-\eta_{l-1})^2}$$

$$=2\sin\frac{\pi}{n}\left\{\left[-a\sin\frac{\pi(2l-1)}{n}+b\cos\frac{\pi(2l-1)}{n}\right]^2+\left[-c\sin\frac{\pi(2l-1)}{n}+d\cos\frac{\pi(2l-1)}{n}\right]^2\right\}^{\frac{1}{2}},$$

因此总长度是

$$L(T)=2\sin\frac{\pi}{n}\sum_{l=1}^{n}\left\{\left[-a\sin\frac{\pi(2l-1)}{n}+b\cos\frac{\pi(2l-1)}{n}\right]^2+\left[-c\sin\frac{\pi(2l-1)}{n}+d\cos\frac{\pi(2l-1)}{n}\right]^2\right\}^{\frac{1}{2}};$$

而新的 $n$ 边形的面积不难算出是

$$A(T)=|ad-bc|\frac{n}{2}\sin\frac{2\pi}{n}.$$

因此本节开始提出的问题的解答已由十一节公式 (6) 给出:

$$\frac{(A(T))^{\frac{1}{2}}}{L(T)}\leqslant\frac{\left(\dfrac{n}{2}\sin\dfrac{2\pi}{n}\right)^{\frac{1}{2}}}{2n\sin\dfrac{\pi}{n}}.$$

当 $n\to\infty$ 时, $n$ 边形趋于圆, $A(T)$ 与 $L(T)$ 各趋于该圆经 $(T)$ 变换而变成的椭圆的面积 $A$ 与周长 $L$, 而且有不等式

$$A\leqslant\frac{1}{4\pi}L^2$$

这式是有名的等边长问题的特例. 但 (5) 式比 (6) 式更精确, 其极限已如上节所述, 因而改进了等边界问题的不等式. 必须指出, 适合原不等式的函数类是很宽的, 对改进了的不等式来说范围狭窄了很多.

我们这儿只不过就平面问题作了一个简单的开端. 所联系到的与格子论, 群论, 不等式论, 变分法等有关的问题还不少呢? 但写得太多是篇幅所不允许的, 并且可能有人会说有些牵强附会了. 实质上, 千丝万缕的关系看来若断若续, 而这正是由此及彼, 由表及里的线索呢! 总之, 想, 联想, 看, 多看, 问题只会愈来愈多的. 至于运用之妙, 那只好存乎其人了! 但习惯于思考联想的人一定会走得深些远些; 没有思考联想的人, 虽然读破万卷书, 依然看不到书外的问题.

1963 年除夕初稿

1964 年 1 月 12 日完稿于铁狮子坟

(据北京出版社 1979 年版排印)

# 第二章

## 创造型工作(I)、探路(II)之代表作

甲. 近似分析中的数论方法……89
● 关于多重积分的近似计算的若干注记…(华罗庚、王元) 91
● 某类函数插值公式的一个注记……(王元) 95
● 丢番图逼近与数值积分(I)……(华罗庚、王元)100
● 丢番图逼近与数值积分(II)……(华罗庚、王元) 104
● 多维周期函数的数值积分……(华罗庚、王元) 108
● 关于一类函数的插入公式……(王元) 123
● 论一致分布与近似分析——数论方法(I)
……(华罗庚、王元) 127
● 论一致分布与近似分析——数论方法(II)
……(华罗庚、王元) 153
● 论一致分布与近似分析——数论方法(III)
……(华罗庚、王元) 174
● Applications of Number Theory to Numerical Analysis
……(华罗庚、王元) 190
乙. 应用统计中的数论方法……421
● 关于均匀分布与试验设计(数论方法)
……(王元、方开泰) 423
● 应用统计中的数论方法(I)……(王元、方开泰) 430
● 应用统计中的数论方法(II)……(王元、方开泰) 448
● 混料均匀设计……(王元、方开泰) 460
● 统计模拟中的数论方法……(王元、方开泰) 472



(甲)

# 近似分析中的数论方法

# 关于多重积分的近似计算的若干注记 *

华罗庚　王　元

(中国科学院数学研究所)

命 $f(x_1,\cdots,x_s)$ 是对每一变数周期皆为 1 的函数, 且可以表为

$$f(x_1,\cdots,x_s)=\sum_{m_1=-\infty}^{\infty}\cdots\sum_{m_s=-\infty}^{\infty}C(m_1,\cdots,m_s)e^{2\pi i(m_1x_1+\cdots+x_sx_s)}, \tag{1}$$

此处

$$C(m_1,\cdots,m_s)=\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_s)e^{-2\pi i(m_1x_1+\cdots+m_sx_x)}\mathrm{d}x_1\cdots\mathrm{d}x_s. \tag{2}$$

给出正整数 $q$ 及正整数系 $1\leqslant a_i<q(1\leqslant i\leqslant q)$, 则得

$$\sum_{t=1}^{q}f\left(\frac{a_1t}{q},\cdots,\frac{a_st}{q}\right)=\sum_{m_1=-\infty}^{\infty}\cdots\sum_{m_s=-\infty}^{\infty}C(m_1,\cdots,m_s)\sum_{t=1}^{q}e^{2\pi i(m_1a_1+\cdots+m_sa_s)t/q}$$

$$=q\sum_{\substack{m_1=-\infty\\ a_1m_1+\cdots+a_sm_s\equiv 0(\bmod\ q)}}^{\infty}\cdots\sum_{m_s=-\infty}^{\infty}C(m_1,\cdots,m_s).$$

由 (2) 即得

$$\left|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_s)dx_1\cdots dx_s-\frac{1}{q}\sum_{t=1}^{q}f\left(\frac{a_1t}{q},\cdots,\frac{a_st}{q}\right)\right|$$

$$=\left|\sum_{a_1m_1+\cdots+a_sm_s\equiv 0(\bmod\ q)}{}' C(m_1,\cdots,m_s)\right|=R, \tag{3}$$

此处 $\Sigma'$ 表示在和中去掉 $m_1=m_2=\cdots=m_s=0$ 的那一项.

现在考虑在以下的条件下关于 $R$ 的估计.

$$|C(m_1,\cdots,m_s)|\leqslant\frac{C}{[(|m_1|+1)\cdots(|m_s|+1)]^{\alpha}}, \tag{4}$$

* 原载《科学记录》新辑第 4 卷第 1 期, 1960 年.

此处 $\alpha>1, C>0$ 均为常数, 易见

$$R\leqslant \sum_{\alpha_1m_1+\cdots+\alpha_sm_s\equiv 0(\bmod\ q)} \frac{1}{[(|m_1|+1)\cdots(|m_s|+1)]^\alpha}$$
$$\leqslant \sideset{}{'}\sum_{\substack{\alpha_1m_1+\cdots+\alpha_sm_s\equiv 0(\bmod\ q)\\ |m_i|\leqslant \frac{1}{2}q}} \frac{1}{[(|m_1|+1)\cdots(|m_s|+1)]^\alpha}+O\left(\frac{1}{q^\alpha}\right).$$

最后一和又等于 $q^{-s\alpha}\Omega$, 此处

$$\Omega=\sideset{}{'}\sum_{\substack{a_1m_1+\cdots+a_sm_s\equiv 0(\bmod\ q)\\ |m_i|\leqslant \frac{q}{2}}} \frac{1}{\left[\left(\left\langle\frac{m_1}{q_1}\right\rangle+\frac{1}{q}\right)\cdots\left(\left\langle\frac{m_s}{q_s}\right\rangle+\frac{1}{q}\right)\right]^\alpha},$$

而其中 $\langle\xi\rangle$ 表示 $\xi$ 与其最近的整数的距离.

Коробов[1] 证明了：当 $q=p$ 为素数时, 存在 $a_1,\cdots,a_s$ 使

$$R=O\left(\frac{\log^{\alpha s}p}{p^\alpha}\right). \tag{5}$$

在此及以下, 与 "$O$" 有关的常数均可以具体写出. 他的工作仅能证明 $a_i$ 是存在的, 故在具体用于计算时, 尚有困难, 他还证明了, 恒存在适合 (1)、(4) 的函数, 对于一切 $q$ 及任何适合 $1\leqslant a_1,\cdots,a_s\leqslant q$ 的诸 $a_i$, 皆有

$$R\geqslant\frac{C}{q^\alpha}. \tag{6}$$

故数值积分的中心问题似乎在于寻求具体的 $a_1,\cdots,a_s,q$. 本文证明了：

**定理 1** 命 $n\geqslant 3$ 为自然数. 又命 $a_1=1, a_2=p_n=\frac{1}{2}\{(1+\sqrt{2})^n+(1-\sqrt{2})^n\}, q=q_n=\frac{1}{2\sqrt{2}}\{(1+\sqrt{2})^n-(1-\sqrt{2})^n\}$, 则

$$R=O\left(\frac{\log q_n}{q_n^\alpha}\right). \tag{7}$$

**证** 如能证明

$$\Omega=2\sum_{1\leqslant x\leqslant\frac{1}{2}q_n}\frac{1}{\left(\left\langle\frac{x}{q_n}\right\rangle\left\langle\frac{q_nx}{q_n}\right\rangle\right)^\alpha}=O(q_n^\alpha\log q_n),$$

即足. 把 $\Omega$ 分为 $n$ 个分和:

$$J_m=\sum_{q_{m-1}\leqslant x<q_m}\frac{1}{\left(\left\langle\frac{x}{q_n}\right\rangle\left\langle\frac{p_n}{q_n}x\right\rangle\right)^\alpha},\quad m=2,\cdots,n,$$

最后一个 $J_m$ 当然只由 $q_{n-1}$ 到 $\frac{1}{2}q_n$.

当 $q_{m-1} \leqslant x < q_m$ 时, 命 $y$ 为适合下面不等式的整数:

$$\left|y - x\frac{p_n}{q_n}\right| \leqslant \frac{1}{2},$$

方程组

$$\begin{cases} x = q_m u + p_m v, \\ y = p_m u + 2q_m v \end{cases}$$

常有整数解, 故

$$q_n y - p_n x = (-1)^m (q_{n-m}u - p_{n-m}v),$$

又显然 $uv < 0$, 且对于区间 $q_{m-1} \leqslant x < q_m$ 中不同的 $x$ 所对应的 $v$ 亦必不同, 故

$$J_m = O\left(\sum_{v\neq 0} \frac{1}{\left(\frac{q_{m-1}}{q_n}\frac{|v|q_{n-m}}{q_n}\right)^{\alpha}}\right) = O(q_n^{\alpha}).$$

因此,

$$R = O(nq_n^{-\alpha}) + O(q_n^{-\alpha}) = O(q_n^{-\alpha}\log q_n).$$

定理证毕.

**附记** 在此仅仅为了简明计, 故用了二次域 $R(\sqrt{2})$, 实际上用任意二次域都是可以的. 特别取 $R(\sqrt{5})$, 即为 Fibonacci 数. 作者认为, 用完实域代替二次域这一方法可能解决高维的问题, 但看来并不是没有困难的.

用同法, 我们改进了 (6), 得到

**定理 2** 恒存在适合 (1), (4) 的函数, 对于一切自然数及任何适合 $1 \leqslant a_1, \cdots, a_s \leqslant q$ 的诸整数 $a_i$, 皆有

$$R \geqslant \frac{C' \log q}{q^{\alpha}}, \tag{8}$$

此处 $C' > 0$ 为常数.

因此, 当 $s = 2$ 时, 我们给出了具体的 $a_i$, 并且得到了最好的结果.

此外, 当 $r > s$ 时, 在条件

$$\left|\frac{\partial^r f(x_1, \cdots, x_s)}{\partial x_{1}^{i_1} \cdots \partial_{s}^{i_s}}\right| \leqslant K(i_1 + \cdots + i_s = r, K \text{ 为常数}) \tag{4'}$$

之下, 我们得到

**定理 3** 命 $q$ 为正整数; $p_1 < \cdots < p_s$ 是区间 $q^{\frac{1}{s}} \leqslant n \leqslant 2^s q^{\frac{1}{s}}$ 中任意 $s$ 个两两

互素的整数, $a_i = \frac{p_1 \cdots p_s}{p_i}$. 则

$$\left|\int_0^1 \cdots \int_0^1 f(x_1, \cdots, x_s)\mathrm{d}x_1 \cdots \mathrm{d}x_s - \int_0^1 f(a_1 x, \cdots, a_s x)\mathrm{d}x\right| = O\left(\frac{1}{q^{\frac{r}{s}}}\right). \quad (9)$$

由此推出

**定理 4** 在定理 3 的假定下, 有

$$\left|\int_0^1 \cdots \int_0^1 f(x_1, \cdots, x_s)\mathrm{d}x_1 \cdots, \mathrm{d}x_s - \frac{1}{q}\sum_{k=1}^{q} f\left(\frac{a_1 k}{q}, \cdots, \frac{a_s k}{q}\right)\right| = O\left(\frac{1}{q^{\frac{r}{s}}}\right). \quad (10)$$

## 参 考 文 献

[1] Корбов, Н. М. 1959, ДАН, ДССР, **124** (6), 1207~1210.

# 某类函数插值公式的一个注记 *

王 元

(中国科学院数学研究所)

命 $E_s^\alpha$ 表示下面函数构成的函数类:

$$f(x_1, \cdots x_s) = \sum \overset{\infty}{\underset{-\infty}{\cdots}} \sum C(m_1, \cdots, m_s) e^{2\pi i(m_1 x_1 + \cdots + m_s x_s)}, \tag{1}$$

此处 Fourier 系数满足

$$|C(m_1, \cdots, m_s)| \leqslant \frac{1}{(\bar{m}_1 \cdots \bar{m}_s)^\alpha}, \tag{2}$$

其中 $\bar{m} = \max(1, |m|)$ 及 $\alpha > 1$ 为一个绝对常数.

命 $N$ 为一个奇素数及 $N_1$ 为满足 $3 < N_1 < \frac{N}{ln^s N}$ 的一个整数, 又命

$$\tilde{C}(m_1, \cdots, m_s) = \frac{1}{N} \sum_{k=1}^{N} f\left(\frac{ka_1}{N}, \cdots, \frac{ka_s}{N}\right) e^{-2\pi i \frac{a_1 m_1 + \cdots + a_s m_s}{N} k}, \tag{3}$$

$$P(x_1, \cdots, x_s) = \sum_{\bar{m}_1 \cdots \bar{m}_s < N_1} \tilde{C}(m_1, \cdots, m_s) e^{2\pi i(m_1 x_1 + \cdots + m_s x_s)}, \tag{4}$$

及

$$\varDelta = \min_a \sup_{f \in E_s^\alpha} \int_0^1 \cdots \int_0^1 |f(x_1, \cdots, x_s) - P(x_1, \cdots, x_s)|^2 \mathrm{d}x_1, \cdots, \mathrm{d}x_s, \tag{5}$$

此处 $a_1, \cdots, a_s$ 为整数.

Rjabenkii[1] 首先证明了

$$\varDelta \leqslant A_0(N_1^{2\alpha+1} N^{-2\alpha} \ln^{2\alpha s+s-1} N + N_1^{-(2\alpha-1)} \ln^{s-1} N_1)^{**} \tag{6}$$

(亦见 [2]).

本文之目的为给出下面的改进:

**定理 1**

$$A_1(N_1^{2\alpha} N^{-2\alpha} + N_1^{-(2\alpha-1)} \ln^{s-1} N_1) \leqslant \varDelta$$

* 原载: 中国科学, **10**, 6, 1961, 632~636.

** 在本文中, $A_\nu$ 与 $B_\nu$ 均表示仅依赖于 $\alpha$、$s$ 的正常数.

$$\leqslant A_2(N_1^{2\alpha}N^{-2\alpha}\ln^{2\alpha(s-1)}N+N_1^{-(2\alpha-1)}\ln^{s-1}N_1). \tag{7}$$

显然这一定理已不允许再有实质的改进. 由此立即推出

**定理 2** 命 $N_1=[N^{\frac{2\alpha}{4\alpha-1}}\ln^{-\frac{(2\alpha-1)(s-1)}{4\alpha-1}}N]$. 则存在整数 $a_i=a_i(N)(1\leqslant i\leqslant s)$ 使不等式

$$\int_0^1\cdots\int_0^1|f(x_1,\cdots,x_s)-P(x_1,\cdots,x_s)|^2\mathrm{d}x_1\cdots\mathrm{d}x_s$$
$$\leqslant A_3N^{-\frac{2\alpha(2\alpha-1)}{4\alpha-1}}\ln^{\frac{4\alpha^2}{4\alpha-1}(s-1)}N \tag{8}$$

对于任何 $f(x_1,\cdots,x_s)\in E_s^\alpha$ 皆成立.

**定理 1 的证明：1)** 不等式 (7) 的右端可以由下面二引理推出：

**引 1** (Bahvalov[3]). 存在 $a_i=a_i(N)(1\leqslant i\leqslant s)$ 使

$$\sum_{a_1l_1+\cdots+a_sl_s\equiv0(\mathrm{mod}\ N)}\overset{\infty}{\underset{-\infty}{\cdots}}\sum{}'\frac{1}{(\bar{l}_1,\cdots,\bar{l}_s)^\alpha}\leqslant A_4\frac{\ln^{\alpha(s-1)}N}{N^\alpha}, \tag{9}$$

此处 $\Sigma'$ 表示一个和, 其中除去 $l_1=\cdots=l_s=0$ 的一项.

**引 2** 命 $l_1,\cdots,l_s$ 为非负整数及 $1<N_1\leqslant\bar{l}_1\cdots\bar{l}_s/3^s$, 则

$$\sum_{\substack{\bar{m}_1\cdots\bar{m}_s<N_1\\ m_j\geqslant0}}\frac{1}{[(\overline{l_1-m_1})\cdots(\overline{l_s-m_s})]^\alpha}\leqslant B_s\frac{N_1^\alpha}{(\bar{l}_1,\cdots,\bar{l}_s)^\alpha}. \tag{10}$$

**证** 当 $s=1$ 时有

$$\sum_{0\leqslant m_1<N_1}\frac{1}{(\overline{l_1-m_1})^\alpha}<\left(\frac{3}{2}\right)^\alpha\frac{N_1}{\bar{l}_1^\alpha}\leqslant B_1\frac{N_1^\alpha}{\bar{l}_1^\alpha}.$$

现在假定引理对于 $s\leqslant k$ 成立, 则

$$\sum_{\substack{\bar{m}_1\cdots\bar{m}_{k+1}<N_1\\ m_j\geqslant0}}\frac{1}{[(\overline{l_1-m_1})\cdots(\overline{l_{k+1}-m_{k+1}})]^\alpha}$$
$$\leqslant\sum\nolimits_1+\sum\nolimits_2+\cdots\sum\nolimits_{k+1},$$

其中

$$\sum\nolimits_i=\sum_{\substack{\bar{m}_1\cdots\bar{m}_{k+1}<N_1\\ m_1<\frac{l_i}{2}\\ m_j\geqslant0}}\frac{1}{[(\overline{l_1-m_1})\cdots(\overline{l_{k+1}-m_{k+1}})]^\alpha}.$$

(i) 假定 $N_1 < \frac{\bar{l}_2\cdots\bar{l}_{k+1}}{3^k}$, 则

$$\sum\nolimits_1 \leqslant \sum_{0\leqslant m_1<\frac{l_1}{2}} \frac{1}{(\overline{l_1-m_1})^\alpha} \sum_{\substack{\bar{m}_2\cdots\bar{m}_{k+1}<\frac{N_1}{\bar{m}_1}\\ m_j\geqslant 0}} \frac{1}{[(\overline{l_2-m_2})\cdots(\overline{l_{k+1}-m_{k+1}})]^\alpha}$$

$$\leqslant B_k\frac{N_1^\alpha}{(\bar{l}_2\cdots\bar{l}_{k+1})^\alpha}\sum_{0\leqslant m_1<\frac{l_1}{2}}\frac{1}{\bar{m}_1^\alpha(\overline{l_1-m_1})^\alpha}$$

$$< 2^{\alpha+1}B_k\zeta(\alpha)\frac{N_1^\alpha}{(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha}.$$

(ii) 假定 $N_1 \geqslant \frac{\bar{l}_2\cdots\bar{l}_{k+1}}{3^k}$, 则

$$\sum\nolimits_1 \leqslant \sum_{0\leqslant m_1<\frac{3^kN_1}{\bar{l}_2\cdots\bar{l}_{k+1}}} \frac{1}{(\overline{l_1-m_1})^\alpha} \sum_{\substack{\bar{m}\cdots\bar{m}_{k+1}<\frac{N_1}{\bar{m}_1}\\ m_j\geqslant 0}} \frac{1}{[(\overline{l_2-m_2})\cdots(l_{k+1}-m_{k+1})]^\alpha}$$

$$+\sum_{\frac{3^kN_1}{\bar{l}_2\cdots\bar{l}_{k+1}}\leqslant m_1<\frac{l_1}{2}} \frac{1}{(\overline{l_1-m_1})^\alpha} \sum_{\substack{\bar{m}_2\cdots\bar{m}_{k+1}<\frac{N_1}{\bar{m}_1}\\ m_j\geqslant 0}} \frac{1}{[(\overline{l_2-m_2})\cdots(\overline{l_{k+1}-m_{k+1}})]^\alpha}$$

$$\leqslant (3\zeta(\alpha))^k\sum_{0\leqslant m_1<\frac{3^kN_1}{\bar{l}_2\cdots\bar{l}_{k+1}}}\frac{1}{(\overline{l_1-m_1})^\alpha}$$

$$+\sum_{\frac{3^kN_1}{\bar{l}_2\cdots\bar{l}_{k+1}}\leqslant m_1<\frac{l_1}{2}}\frac{1}{(\overline{l_1-m_1})^\alpha}\cdot\frac{B_kN_1^\alpha}{\bar{m}_1^\alpha(\bar{l}_2\cdots\bar{l}_{k+1})^\alpha}$$

$$\leqslant 3^{\alpha+k}\zeta^k(\alpha)\frac{3^kN_1}{\bar{l}_2\cdots\bar{l}_{k+1}\bar{l}_1^\alpha}+\frac{2^\alpha B_kN_1^\alpha\zeta(\alpha)}{(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha}$$

$$\leqslant (3^{\alpha+k+\alpha k}\zeta^k(\alpha)+2^\alpha B_k\zeta(\alpha))\frac{N_1^\alpha}{(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha}.$$

所以

$$\sum\nolimits_1 \leqslant \left(\frac{B_{k+1}}{k+1}\right)\frac{N_1^\alpha}{(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha}.$$

类似地

$$\sum\nolimits_i \leqslant \left(\frac{B_{k+1}}{k+1}\right)\frac{N_1^\alpha}{(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha},\quad (2\leqslant i\leqslant k+1).$$

因此

$$\sum_{\substack{\bar{m}_1\cdots\bar{m}_{k+1}<N_1\\ m_j\geqslant 0}}\frac{1}{[(\overline{l_1-m_1})\cdots(\overline{l_{k+1}-m_{k+1}})]^\alpha}\leqslant B_{k+1}\frac{N_1^\alpha}{(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha}.$$

从而由归纳法即得引理.

2) 取

$$f(x_1,\cdots,x_s)=\sum\overset{\infty}{\underset{-\infty}{\cdots}}\sum\frac{1}{(\bar m_1\cdots\bar m_s)^\alpha}e^{2\pi i(m_1x_1+\cdots+m_sx_s)}.$$

则对于任何 $a_1,\cdots,a_s$, 我们有

$$\int_0^1\cdots\int_0^1|f(x_1,\cdots,x_s)-P(x_1,\cdots,x_s)|^2\mathrm{d}x_1\cdots\mathrm{d}x_s$$

$$=\sum_{\bar m_1\cdots\bar m_s<N_1}\left(\sum_{\alpha_1l_1+\cdots+a_sl_s\equiv0(\bmod\ N)}\overset{\infty}{\underset{-\infty}{\cdots}}\sum{}'\frac{1}{[(\overline{l_1-m_1})\cdots(\overline{l_{k+1}-m_{k+1}})]^\alpha}\right)^2$$

$$+\sum_{\bar m_1\cdots\bar m_s\geqslant N_1}\frac{1}{(\bar m_1\cdots\bar m_s)^{2\alpha}}$$

$$\geqslant\sum_{m_1m_2<N_1}\sum_{\alpha_1l_1+\alpha_2l_2\equiv0(\bmod\ N)}{}'\frac{1}{[(l_1-m_1)(l_2-m_2)]^{2\alpha}}$$

$$+A_5N_l^{-(2\alpha-1)}\ln^{s-1}N_1.$$

首先假定 $(a_i,N)=1(i=1,2)$. 命同余式 $a_1x\equiv-a_2(\bmod\ N)$的解为 $x=a(|a|<N)$ 及 $p_t/q_t$ 为 $a/N$ 的 $t$ 次渐近分数, $\frac{p_n}{q_n}=\frac{a}{N}$ 及 $1=q_0<q_1<\cdots<q_r<N_1\leqslant q_{r+1}<\cdots<q_n=N$. 由于对于任何整数 $b$ 皆有

$$|p_nq_t-q_nb|\leqslant\frac{q_n}{q_{t+1}},\quad(0\leqslant t\leqslant n-1),$$

所以

$$\int_0^1\cdots\int_0^1|f(x_1,\cdots,x_s)-P(x_1,\cdots,x_s)|^2\mathrm{d}x_1\cdots\mathrm{d}x_s$$

$$\geqslant\sum_{m_1m_2<N_1}\sum_{l_1\equiv al_2(\bmod\ N)}\frac{1}{[(\overline{l_1-m_1})(\overline{l_2-m_2})]^{2\alpha}}+A_5N_l^{-(2\alpha-1)}\ln^{s-1}N_1.$$

$$\geqslant\sum_{m_1m_2<N_1}\sum_{t=0}^{n-1}\frac{1}{\left[(\overline{q_t-m_1})\left(\overline{\left[\frac{q_n}{q_{t+1}}\right]-m_2}\right)\right]^{2\alpha}}$$

$$+A_5N_1^{-(2\alpha-1)}\ln^{s-1}N_1$$

$$\geqslant\left(\frac{q_{r+1}}{q_n}\right)^{2\alpha}+A_5N_1^{-(2\alpha-1)}\ln^{s-1}N_1$$

$$\geqslant A_2(N_1^{2\alpha}N^{-2\alpha}+N_1^{-2\alpha-1}\ln^{s-1}N_1).$$

显然对于情况 $(a_1, N) > 1$ 或 $(a_2, N) > 1$, 上面的估计仍成立, 故得不等式 (7) 之右端.

类似地, 我们有

**定理 3** 命 $N_1 = [N^{\frac{\alpha}{2\alpha-1}} \ln^{\frac{(s-1)(1-\alpha)}{2\alpha-1}} N]$. 则存在 $a_i = a_i(N)(1 \leqslant i \leqslant s)$ 使

$$\left| f(x_1, \cdots, x_s) - \sum_{\bar{m}_1 \cdots \bar{m}_2 < N_1} \left[ \frac{1}{N} \sum_{k=1}^{N} f\left( \frac{\alpha_1 k}{N}, \cdots, \frac{\alpha_s k}{N} \right) e^{-2\pi i \frac{(m_1 \alpha_1 + \cdots + m_s \alpha_s)}{N} k} \right] e^{2\pi i (m_1 x_1 + \cdots + m_s x_s)} \right| \leqslant A_6 N^{-\frac{\alpha(\alpha-1)}{2\alpha-1}} \ln^{\frac{\alpha^2}{2\alpha-1}(s-1)} N \quad (11)$$

对于所有 $f(x_1, \cdots, x_s) \in E_s^\alpha$ 皆成立.

**附记** 用本文的方法, 我们可以改进 Sahov 一个定理, 即将他定理中的误差项 $O(N^{-\frac{\alpha-1}{2}+\varepsilon})$ 改进为 $O(N^{-\alpha/2+\varepsilon})$.

## 参 考 文 献


[1] V. S. Rjabenkii, Tables and interpolation of a certain class of functions, Dokl. Akad. Nauk SSSR, **131**, 1960, 1025~1027.

[2] S. A. Smoljak, Interpolation and quadrature formulas for the Classes $W_s^\alpha$ and $E_s^\alpha$, Dokl. Akad, Nauk SSSR, **131**, 1960, 1028~1031.

[3] N. S. Bahvalov, Approximate computation of multiple integrals, Vestnik Moskow Uniw; Ser. Mat. Meh. Astr. Fiz. Him; **4**, 1959, 3~18.

[4] Yu. N. Sahov, On the approximate solution of Volterra equation of second type by the method of iteration, Dokl. Akad. Nauk SSSR, **128**, 1959, 1136~1139.

# 丢番图逼近与数值积分 (I)*

华罗庚　　王　元
(中国科学技术大学)　(中国科学院数学研究所)

I. 通常计算多重积分的 Monte carlo 方法是经验性质的, 这是无法估计正确误差的方法. H. M. Коробов运用数论方法, 在 1959 年引进了 “极值系数法”, 得出了一个可以精密估计误差的方法. 这一方法就是用预先选定的最佳分布上的函数值的算术平均来逼近多重积分. 但在实际应用这一方法时, 估计误差的性质基本上是 “从所有可能的分布中选取最佳的分布”[1], 即算出各种分布所对应的数值积分的误差, 而确定出最佳分布, 因此必须经过冗长的计算. 在本文中, 我们运用代数数论及丢番图逼近论直接提出了两种分布, 能够得到与Коробов方法具有同样精密的有效数字的结果 [2]. 在下面的一个例子 (例 2) 里可以看到, 他用一亿次四则运算所能得到的结果, 我们只要用五万次运算就可以得到有同样精密的有效数字的结果. 这个例子只是为了可以与Коробов的结果比较而提出的. 实质上, 随着对精密度要求的提高, 运算次数的相差更为悬殊, 结果不是例 2 中所提的 2 000 倍, 而是更大的与计算量俱增的倍数. 结果因而涉及计算机的大小问题. 换言之, 在他们不能实现的时候, 我们的方法还有余地. 即使大到了连我们也无法估计误差的时候, 但我们还能建议实际计算数值积分的方法, 这个优点是他们所没有的. 详细言之, 欲得到 $N$ 个分点的分布位置及由此而得到的数值积分的误差, Коробов方法须经过 $O(N^2)$ 次四则运算, 当 $N$ 较大时, 可以减低至 $O(N^{4/3})$ 次, 而且必须经过这些运算后, 才能得到 $N$ 个点的最佳分布. 而我们的方法基本上不需要经过计算即得到分点分布的位置 (即是说可以在电动计算机上实现, 而且量不大的计算). 而欲得到由这些分点来计算积分的误差, 所需的计算量亦仅为 $O(N)$ 次运算. 另一方面, 用我们的方法在电子计算机上进行计算时, 程序亦比较简单 (不需要将许多数进行比较). 因此本文指出了当积分维数较高或分点较多时, 精密计算多重积分的途径. 同时, 本文还将给出多个无理数联立有理逼近的具体模型, 因此在理论方面, 亦指出了一系列值得进一步探讨的问题.

II. 命 $f(x_1,\cdots,x_s)$ 有绝对收敛的 Fourier 展开

$$f(x_1,\cdots,x_s)=\sum \mathop{\cdots}\limits_{-\infty}^{\infty} \sum C(m_1,\cdots,m_s)e^{2\pi i(m_1x_1+\cdots+m_sx_s)},$$

* 原载:《科学通报》, 1964 年 6 月.

其系数适合于

$$|C(m_1,\cdots,m_s)| \leqslant \frac{1}{(\bar{m}_1\cdots\bar{m}_s)^\alpha},$$

其中 $\bar{m}=\max(1,|m|),\alpha>1$ 为常数. 这种函数的全体记为 $E_s^\alpha$. 当 $\alpha=2$ 时, 对于整数组 $q;a_1,\cdots,a_s$ 有

$$\sup_{f\in E_s^2}\left|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_s)\mathrm{d}x_1\cdots\mathrm{d}x_s-\frac{1}{q}\sum_{k=1}^{q} f\left(\frac{a_1k}{q},\cdots,\frac{a_s,k}{q}\right)\right|$$
$$\leqslant\left(\frac{\pi^2}{6}\right)^s H(q;a_1,\cdots,a_s),$$

此处

$$H(q;a_1,\cdots,a_s)$$
$$=\begin{cases}\dfrac{3^s}{q}\left[1+2\displaystyle\sum_{k=1}^{\frac{q-1}{2}}\prod_{v=1}^{s}\left(1-2\left\{\frac{a_vk}{q}\right\}\right)^2\right]-1, & 若\ 2\nmid q,\\ \dfrac{3^s}{q}\left[1+\displaystyle\prod_{v=1}^{s}\left(1-2\left\{\frac{a_v}{2}\right\}\right)^2+2\sum_{k=1}^{\frac{q}{2}-1}\prod_{v=1}^{s}\left(1-2\left\{\frac{a_vk}{q}\right\}\right)^2\right]-1, & 若\ 2|q.\end{cases}$$

数值积分的中心问题在于有效地定出 $q;a_1,a_2,\cdots,a_s$ 使 $H(q;a_1,\cdots,a_s)$ 较小. Коробов[1] 证明当 $q=p$ 为素数时, 存在整数 $a$ 使

$$H(p;1,a\cdots,a^{s-1})=H(p,a)\ll\frac{\log^{2s}p}{p^2}.$$

对于 $p$, 欲求出使 $H(p,a)$ 取极小值的 $a$, 则必须经过 $O(p^2)$ 次加减乘除的运算, 经过改进后, 计算量可以减低至 $O(q^{4/3})$.

这两文的目的在于提供两个直接给出 $q;a_1,\cdots,a_s$ 的实用方法. 就已有的数值结果而言, 与Коробов方法具有同样精密的有效数字 [2].

命 $p_1,\cdots,p_t$ 为 $t$ 个不同的素数及

$$\varepsilon_1,\cdots,\varepsilon_{2^t-1}$$

为代数数域 $R(\sqrt{p_1},\cdots,\sqrt{p_t})$ 的一组独立单位. 即诸 Pell 氏方程

$$x^2-p_{i_1}\cdots p_{i_k}y^2=\pm4.\quad (x>0,y>0\ 为整数)$$

的最小解 $\dfrac{x}{2}+\dfrac{\sqrt{p_{i_1}\cdots p_{i_k}}y}{2}$, 此处 $k\geqslant 1,1\leqslant i_1<\cdots<i_k\leqslant t$ 为 $1,2,\cdots,t$ 的任意

选取. 我们可以取

$$\varepsilon_i=\begin{cases}\dfrac{x_i}{2}+\dfrac{\sqrt{d_i}y_i}{2}, & x_i\equiv y_i\equiv 1(\bmod 2)\quad(1\leqslant i\leqslant m),\\ x_i+\sqrt{d_i}y_i & (m+1\leqslant i\leqslant 2^t-1).\end{cases}$$

取正整数 $n_1,\cdots,n_t$ 使对于任意 $i\neq j$ 皆有

$$c_1\varepsilon_{j^j}^{n}<\varepsilon_{i^i}^{n}<c_2\varepsilon_{j^j}^{n},$$

此处 $c_1,c_2$ 为正常数. 考虑展开式

$$\varepsilon_{1^1}^{n}\cdots\varepsilon_{2^t-1}^{n_{2^2-1}}=q_0^{(n)}+q_1^{(n)}\varepsilon_1+\cdots+q_m^{(n)}\varepsilon_m+q_{m+1}^{(n)}\sqrt{d_{m+1}}+\cdots+q_{2^t-1}^{(n)}\sqrt{d_{2^t-1}}\quad(1)$$

及 $2^t-1$ 个变换

$$(\sigma_{i_1\cdots i_k})\sqrt{p_v}\to\begin{cases}\sqrt{p_v}, & 若\ v\neq i_j(1\leqslant j\leqslant k),\\ -\sqrt{p_v}, & 若\ v=i_j(1\leqslant j\leqslant k).\end{cases}\quad(2)$$

将变换 (2) 作用于方程 (1), 则得 $2^t-1$ 个方程. 与 (1) 联立, 解之得

$$\begin{aligned}&2q_0^{(n)}+q_1^{(n)}+\cdots+q_m^{(n)}\\&=\varepsilon_{1^1}^{n}\cdots\varepsilon_{2^t-1}^{n_{2^t-1}}/2^{t-1}+O(|\varepsilon_1|^{-n_1}),\\ q_i^{(n)}&=\varepsilon_{1^1}^{n}\cdots\varepsilon_{2^t-1}^{n_{2^t-1}}/2^{t-1}\sqrt{d_i}+O(|\varepsilon_1|^{-n_1})\quad(1\leqslant i\leqslant m),\\ q_i^{(n)}&=\varepsilon_{1^1}^{n}\cdots\varepsilon_{2^t-1}^{n_{2^t-1}}/2^{t}\sqrt{d_i}+O(|\varepsilon|^{-n_1})\quad(m+1\leqslant i\leqslant 2^2-1).\end{aligned}\quad(3)$$

当 $2|q_1^{(n)},\cdots,q_m^{(n)}$ 时, 定义 $q=q_0^{(n)}+\frac{1}{2}(q_1^{(n)}+\cdots+q_m^{(n)}),a_1=1,a_{i+1}\equiv d_iq_i^{(n)}/2(\bmod q)(1\leqslant i\leqslant m),a_{i+1}\equiv d_iq_i^{(n)}(\bmod d)(m+1\leqslant i\leqslant 2^t-1)$.

当 $2\nmid(q_1^{(n)},\cdots,q_m^{(n)})$ 时, 定义 $q=2q_0^{(n)}+(q_1^{(n)}+\cdots+q_m^{(n)}),a_1=1,a_{i+1}\equiv d_iq_i^{(n)}(\bmod q)(1\leqslant i\leqslant m),a_{i+1}\equiv 2d_iq_i^{(m)}(\bmod q)(m+1\leqslant i\leqslant 2^t-1)$.

由 (3) 可得以下的联立丢番图逼近式

$$\left|\frac{a_i}{q}-\sqrt{d_{i-1}}\right|=O\left(\frac{1}{q^{1+\frac{1}{2^t-1}}}\right)\quad(2\leqslant i\leqslant 2^t).$$

我们就取这样的 $q;a_1=1,a_2,\cdots,a_{2^t}$ 来计算 $H(q;a_1,\cdots,a_{2^t})$.

**III. 例 1** 域 $R(\sqrt{2})$ 有基本单位 $\varepsilon=1+\sqrt{2}$. 若 $\varepsilon^n=q_0^{(n)}+\sqrt{2}q_1^{(n)}$, 则

$$H(q_0^{(n)},1,2q_1^{(n)})\ll\frac{\log q_0^{(n)}}{q_0^{(n)2}}$$

([3] 中已经证明这是最优的结果).

**例 2**　取 $R(\sqrt{2},\sqrt{5})$ 及 $\varepsilon_1=\dfrac{1+\sqrt{5}}{2},\varepsilon_2=1+\sqrt{2},\varepsilon_3=3+\sqrt{10}$. 由展开式

$$\begin{aligned}\varepsilon_1^6\varepsilon_2^4\varepsilon_3^2 &= (9+4\sqrt{5})(17+12\sqrt{2})(19+6\sqrt{10})\\ &= 5787+4092\sqrt{2}+2588\sqrt{5}+1830\sqrt{10}\end{aligned}$$

得出

$$q=5787,\quad a_1=1,\quad a_2=2397,\quad a_3=1366,\quad a_4=939.$$

经实际计算得

$$H(5787;1,2397,1366,939)\leqslant 0.001498.$$

**附记**　建议一个计算 $2^t-1$ 重积分的方法

$$\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_s)\mathrm{d}x_1\cdots\mathrm{d}x_s\approx\frac{\displaystyle\sum_{j=-q^l}^{q^l}\mu_{q,l,j}f(\varepsilon_1 j,\cdots,\varepsilon_{2^t-1}j)}{(2q+1)^l}$$

此处 $f\in E_{2^t-1}^{\alpha}$, $l$ 为 $\geqslant\alpha$ 的最小整数, $\mu_{q,l,j}$ 由展开式

$$\left(\sum_{j=-q}^{q}z^j\right)^l=\sum_{j=-q^l}^{q^l}\mu_{q,l,j}z^j$$

定义, 而 $\varepsilon_1,\cdots,\varepsilon_{2^t-1}$ 为 $R(\sqrt{p_1},\cdots,\sqrt{p_t})$ 的一组独立单位 [4].

志谢: 作者对徐钟济教授的帮助, 致以衷心的感谢. 又承蒙方庆萱同志代为用电子计算机算出本文诸例, 亦致谢意.

## 参 考 文 献

[1] Коробов Н. М., ДАН СССР, **132**, 5, 1009~1012 (1960).
[2] Салтыков А. И., Жур. Выч. Мат, *и МатФиз*., **3**, 1, 181~186 (1963).
[3] 华罗庚、王元, 科学记录新辑, **4**, 1, 4~8 (1960).
[4] Бахвалов Н. С., Вес. МГУ., **4**, 3~18 (1959).

# 丢番图逼近与数值积分 (II)*

华罗庚　　王　元
(中国科学技术大学)　(中国科学院数学研究所)

I. 延用前文的记号. 命 $p \geqslant 5$ 为素数及 $r = \frac{p-3}{2}$. 取代数数域 $R\left(\cos\frac{2\pi}{p}\right)$ 的 $r+1$ 个单位

$$2\cos\frac{2\pi}{p}, 2\cos\frac{4\pi}{p}, \cdots, 2\cos\frac{2(r+1)\pi}{p}, \tag{1}$$

将它们按绝对值的大小排列如下:

$$|\varepsilon_1^{(1)}| > |\varepsilon_2^{(1)}| > \cdots > |\varepsilon_{r+1}^{(1)}|.$$

当 $1 \leqslant v \leqslant r+1$ 时, 引入变换

$$(\sigma_\nu)\ 2\cos\frac{2\pi\mu}{p} \to 2\cos\frac{2\pi\mu\nu}{p} \quad (1 \leqslant \mu \leqslant r+1),$$

这是将集合 (1) 变为自身的变换. 记为

$$(\sigma_v) \quad \varepsilon_\mu^{(1)} \to \varepsilon_\mu^{(\nu)} \quad (1 \leqslant \mu \leqslant r+1).$$

命

$$\varDelta = \begin{pmatrix} 1, & \varepsilon_1^{(1)}, & \cdots, & \varepsilon_r^{(1)} \\ 1, & \varepsilon_1^{(2)}, & \cdots, & \varepsilon_r^{(2)} \\ & & \cdots\cdots & \\ 1, & \varepsilon_1^{(r+1)}, & \cdots, & \varepsilon_r^{(r+1)} \end{pmatrix},$$

$$\tilde{\varDelta} = \frac{1}{p}\begin{pmatrix} 2-\varepsilon_{r+1}^{(1)}, & 2-\varepsilon_{r+1}^{(2)}, & \cdots, & 2-\varepsilon_{r+1}^{(r+1)} \\ \varepsilon_1^{(1)}-\varepsilon_{r+1}^{(1)}, & \varepsilon_1^{(2)}-\varepsilon_{r+1}^{(2)}, & \cdots, & \varepsilon_1^{(r+1)}-\varepsilon_{r+1}^{(r+1)} \\ & & \cdots\cdots & \\ \varepsilon_r^{(1)}-\varepsilon_{r+1}^{(1)}, & \varepsilon_r^{(2)}-\varepsilon_{r+1}^{(2)}, & \cdots, & \varepsilon_r^{(r+1)}-\varepsilon_{r+1}^{(r+1)} \end{pmatrix}$$

则得次之诸性质:

* 原载:《科学通报》, 1964 年 6 月.

(i) $\sum_{\mu=1}^{r+1} \varepsilon_\mu^{(\nu)} = -1$,

(ii) $\prod_{\mu=1}^{r+1} \varepsilon_\mu^{(\nu)} = (-1)^{[\frac{n+2}{2}]}$,

(iii) $|\det\ \Delta| = p^{\frac{r}{2}}$,

(iv) $\Delta\tilde{\Delta} = I$.

命

$$A = \begin{pmatrix} \log\left|\varepsilon_1^{(2)}\right|, & \cdots, & \log\left|\varepsilon_r^{(2)}\right| \\ & \cdots\cdots & \\ \log\left|\varepsilon_1^{(r+1)}\right|, & \cdots, & \log\left|\varepsilon_r^{(r+1)}\right| \end{pmatrix},$$

则得

$$\det\ A \neq 0^*. \tag{2}$$

我们可以取

$$\det\begin{pmatrix} \log\left|\frac{\varepsilon_1^{(3)}}{\varepsilon_1^{(2)}}\right|, & \cdots, & \log\left|\frac{\varepsilon_{r-1}^{(3)}}{\varepsilon_{r-1}^{(2)}}\right| \\ & \cdots\cdots & \\ \log\left|\frac{\varepsilon_1^{(r+1)}}{\varepsilon_1^{(2)}}\right|, & \cdots, & \log\left|\frac{\varepsilon_{r-1}^{(r+1)}}{\varepsilon_{r-1}^{(2)}}\right| \end{pmatrix} \neq 0,$$

从而方程组

$$|\varepsilon_1^{(i)^{a_1}} \cdots \varepsilon_r^{(i)^{a_r}}| = |\varepsilon_1^{(2)^{a_1}} \cdots \varepsilon_r^{(2)^{a_r}}| \quad (3 \leqslant i \leqslant r+1)$$

有唯一的非零解

$$\left(\frac{a_1}{a_r}, \cdots, \frac{a_{r-1}}{a_r}\right).$$

由 (2) 可知, 我们可以假定

$$|\varepsilon_1^{(2)^{a_1}} \cdots \varepsilon_r^{(2)^{a_r}}| < 1.$$

取正整数 $n_r$ 充分大及整数组 $n_1, \cdots, n_{r-1}$ 使

$$\left|\frac{n_i}{n_r} - \frac{a_i}{a_r}\right| < \frac{1}{n} \quad (1 \leqslant i \leqslant r-1),$$

---

* 当 $\det\ A = 0$ 时, 我们可以用下面的独立单位

$$\eta_l = \sin\frac{\pi}{p}g^{l+1} / \sin\frac{\pi}{p}g^l \quad (l = 0, 1, \cdots, r-1),$$

此处 $g$ 为 mod $p$ 的最小原根.

及

$$|\varepsilon_1^{(i)^{n}1} \cdots \varepsilon_r^{(i)^{n}r}| < 1 \quad (2 \leqslant i \leqslant r+1).$$

因此

$$|\varepsilon_1^{(i)^{n}1} \cdots \varepsilon_r^{(i)^{n}r}| = O(|\varepsilon_1^{(1)^{n}1} \cdots \varepsilon_r^{(1)^{n}r}|^{-\frac{1}{r}}) \quad (2 \leqslant i \leqslant r+1).$$

解方程组

$$\varepsilon_1^{(i)^{n}1} \cdots \varepsilon_r^{(i)^{n}r} = q_0^n(2-\varepsilon_{r+1}^{(i)})/p + \sum_{j=1}^{r} q_j^n(\varepsilon_j^{(i)} - \varepsilon_{r+1}^{(i)})/p \quad (1 \leqslant i \leqslant r+1),$$

得出

$$\begin{aligned} q_0^{(n)} &= \varepsilon_1^{(1)^{n}1} \cdots \varepsilon_r^{(1)^{n}r} + O(|\varepsilon_1^{(1)^{n}1} \cdots \varepsilon_r^{(1)^{n}r}|^{-\frac{1}{r}}), \\ q_i^{(n)} &= \varepsilon_1^{(i)^{n}1} \cdots \varepsilon_r^{(1)^{n}r}\varepsilon_i^{(1)} + O(|\varepsilon_1^{(1)^{n}1} \cdots \varepsilon_r^{(i)^{n}r}|^{-\frac{1}{r}}) \quad (1 \leqslant i \leqslant r). \end{aligned} \tag{3}$$

命

$$q = |q_0^{(n)}|, \quad a_1 = 1, \quad a_{i+1} \equiv |q_i^{(n)}| (\bmod\ q) \quad (1 \leqslant i \leqslant r)$$

由 (3) 得如下的联立丢番图逼近式

$$\left|\frac{a_i}{q} - \varepsilon_{i-1}^{(1)}\right| \ll \frac{1}{q^{1+\frac{1}{r}}} \quad (2 \leqslant i \leqslant r+1).$$

我们就取这样的 $q; a_1 = 1, a_2, \cdots, a_{r+1}$ 来计算 $H(q, a_1, \cdots, a_{r+1})$.

诸 $q_i^{(n)} (0 \leqslant i \leqslant r)$ 的实际算法如下：易得展开式

$$\varepsilon_1^{(1)^{n}1} \cdots \varepsilon_r^{(1)^{n}r} = h_1^{(n)}\varepsilon_1^{(1)} + \cdots + h_{r+1}^{(n)}\varepsilon_{r+1}^{(1)},$$

从而

$$q_0^{(n)} = -\sum_{i=1}^{r+1} h_i^{(n)},$$

$$q_i^{(n)} = ph_i^{(n)} - 2\sum_{j=1}^{r+1} h_j^{(n)} \quad (1 \leqslant i \leqslant r).$$

**II. 例 1** 取 $R\left(\cos\frac{2\pi}{5}\right)$ 及 $\varepsilon_1 = 2\cos\frac{4\pi}{5}, \varepsilon_2 = 2\cos\frac{2\pi}{5}$. 由 $2-\varepsilon_2 = \sqrt{5}\varepsilon_2$, $\varepsilon_1 - \varepsilon_2 = -\sqrt{5}$ 及展开式 $\varepsilon_1^n = q_0^{(n)}\varepsilon_2 - q_1^{(n)}$ 得

$$H(|q_0^{(n)}|; 1, |q_1^{(n)}|) \ll \frac{\log|q_0^{(n)}|}{|q_0^{(n)}|^2}. \quad (\text{见}[1])$$

**例 2** 取 $R\left(\cos\frac{2\pi}{7}\right)$ 及

$$\varepsilon_1 = 2\cos\frac{6\pi}{7}, \quad \varepsilon_2 = 2\cos\frac{2\pi}{7}, \quad \varepsilon_3 = 2\cos\frac{4\pi}{7}.$$

由方程 $|\varepsilon_2^\alpha\varepsilon_3^\beta| = |\varepsilon_3^\alpha\varepsilon_1^\beta|$ 得出 $\frac{\alpha}{\beta} \approx 1.356\cdots \approx \frac{4}{3}$.
由展开

$$\varepsilon_1^8\varepsilon_2^6 = -227\varepsilon_1 - 45\varepsilon_2 - 146\varepsilon_3$$

得出

$$q = 418; \quad a_1 = 1, \quad a_2 = 335, \quad a_3 = 103.$$

由实际计算得

$$H(418; 1, 335, 103) \leqslant 0.0108146.$$

**例 3** 分别考虑 $R\left(\cos\frac{2\pi}{11}\right)$ 及 $R\left(\cos\frac{2\pi}{13}\right)$ 得

$$H(9389; 1, 8628, 6408, 2908, 7800) \leqslant 0.0081175,$$
$$H(41204; 1, 38810, 31766, 20480, 5610, 29223) \leqslant 0.0094250.$$

**附记** 建议一个计算重积分的方法:

$$\int_0^1 \cdots \int_0^1 f(x_1, \cdots, x_r)\mathrm{d}x_1 \cdots \mathrm{d}x_r$$
$$\approx \frac{1}{(2q+1)^l} \sum_{j=-q^l}^{q^l} \mu_{q,l,j} f(\varepsilon_1^{(1)} j, \cdots, \varepsilon_r^{(1)} j).$$

## 参 考 文 献

[1] 华罗庚, 王元, 科学记录新辑, 4, 1, 4~8 (1960).

# 多维周期函数的数值积分 *

华罗庚　　　　王　元

(中国科学院数学研究所)

## 提　　要

用代数数论的方法, 本文建议了一个计算多重积分的方法. 文中给出的一个十一维积分的数值结果表明了这一方法的优越性.

## §1. 序　　言

Monte Carla 方法是用单和逼近 $s$ 维重积分的方法. 具体地说, 命 $f(x_1, \cdots, x_s)$ 是一个 $s$ 维周期函数, 假定它的每一变数的周期都是 $1.0 \leqslant x_\nu \leqslant 1 \leqslant (\nu = 1, \cdots, s)$ 称为 $s$ 维方. 假定有一均匀分布的随机变量

$$(x_1^{(i)}, \cdots, x_s^{(i)}), \quad i = 1, 2, \cdots, \tag{1.1}$$

则以平均

$$\frac{1}{N}\sum_{i=1}^{N} f(x_1^{(i)}, \cdots, x_s^{(i)}) \tag{1.2}$$

代积分

$$\int_0^1 \cdots \int_0^1 f(x_1, \cdots, x_s)\mathrm{d}x_1 \cdots \mathrm{d}x_s \tag{1.3}$$

的概率误差是 $O\left(\frac{1}{\sqrt{N}}\right)$.

数论方法是构造一个数列 (1.1), 使 (1.2) 逼近 (1.3) 有绝对误差. 关于这方面较理论的结果已经搜集在我们的小册子之中 [2], 这儿不准备加以论述. 但值得一提的是Коробов[5] 及 Halton[6] 都做出了重要的贡献.

我们的着眼点在于具体的计算方法及数值结果. 最初Коробов用解析数论中的完整三角和法 (见华罗庚 [1]), 找到了一个用单和逼近多重积分的方法, 逼近的绝对误差与 Monte Carlo 方法所得到的概率误差一样, 也就是 $O\left(\frac{1}{\sqrt{N}}\right)$, 这儿 $N$ 是所

* 原载:《中国科学技术大学学报》第 2 卷第 1 期, 1966 年 2 月. 参见 "on Numerical integration of periodic functions of several variables", Sci, Sin; 14, 1965.

用的分点数. 其后他又证明了一条精密度更高的定理. 但这是一条类似存在性的定理. 如欲用于实际计算, 则需要在某种类型的分点中进行全部比较, 求出在这许多种分点中使积分误差最小的一种.

本文用代数数论与丢番图逼近论方法, 建议了一组分点. 而这种分点就与Коробов最佳的分点本质上差不多 (在具体的计算例子中, 有效数字一样). 就这样大大地降低了计算量. 切实些说, 是 $O(P)$ 与 $O(P^{1/2})$ 之比 (或 $O(P)$ 与 $O(P^{3/4})$ 的比). 这当然给我们提供了在同样快速的电子计算机, 可以算出维数更高、精密度更大的结果来的可能性.

为了阐明这一方法的优越性, 我们给出下面的数值结果:

命 $f(x_1,\cdots,x_{11})$ 为一个适合下面条件的周期函数. 则得

$$\left|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_{11})\mathrm{d}x_1\cdots\mathrm{d}x_{11}-\frac{1}{q}\sum_{t=1}^{q}f\left(\frac{a_1t}{q},\cdots,\frac{a_{11}t}{q}\right)\right|$$
$$\leqslant c\left(\frac{\pi^2}{6}\right)^{11}\times 0.2333543,$$

此处

$$\begin{aligned}
q &= 698047, & a_1 &= 1, & a_2 &= 685041,\\
a_3 &= 646274, & a_4 &= 582461, & a_5 &= 494796,\\
a_6 &= 384914, & a_7 &= 254860, & a_8 &= 107051,\\
a_9 &= 642292, & a_{10} &= 467527, & a_{11} &= 284044.
\end{aligned}$$

关于 5 维与 6 维积分的结果将在文后给出.

值得一提的是本文所得到的分点可能用于概率论.

作者对徐钟济教授的帮助, 致以衷心的感谢, 又承蒙万庆萱同志代为用电子计算机算出本文诸例, 亦致谢意.

## §2. 重积分与单和

首先让我们构造一些辅助函数:

**引 2.1** 命 $\{x\}$ 表示 $x$ 的分数部分. 则

$$3(1-2\{x\})^2=\sum_{m=-\infty}^{\infty}\frac{e^{2\pi imx}}{\frac{\pi^2}{6}\bar{m}^2},$$

此处 $\bar{n}=\max(1,|n|)$.

**证**　我们只要证明

$$C_m=\int_0^1(3(1-2\{x\})^2e^{-2\pi imx}\mathrm{d}x=\begin{cases}1, & 若\ m=0,\\ \dfrac{6}{\pi^2m^2}, & 若\ m\neq 0\end{cases}$$

即足矣. 显然有

$$C_m=2\int_0^{1/2}3(1-2x)^2\cos 2\pi mx\mathrm{d}x.$$

当 $m=0$ 时

$$C_0=2\int_0^{1/2}3(1-2x)^2\mathrm{d}x=-(1-2x)^3\Big|_0^{1/2}=1.$$

当 $m\neq 0$ 时, 用两次分部积分得

$$\begin{aligned}C_m&=6(1-2x)^2\frac{\sin 2\pi mx}{2\pi m}\Big|_0^{1/2}+24\int_0^{1/2}\frac{(1-2x)\sin 2\pi mx}{2\pi m}\mathrm{d}x\\&=-\frac{12}{\pi m}(1-2x)\frac{\cos 2\pi mx}{2\pi m}\Big|_0^{1/2}-48\int_0^{1/2}\frac{\cos 2\pi mx}{(2\pi m)^2}\mathrm{d}x=\frac{6}{\pi^2m^2}.\end{aligned}$$

由此可得

**引 2.2**

$$\begin{aligned}g(x_1,\cdots,x_s)&=3^s\prod_{v=1}^{s}(1-2\{x_v\})^2\\&=\sum\overset{\infty}{\underset{-\infty}{\cdots}}\sum\frac{e^{2\pi i(m_1x_1+\cdots+m_sx_s)}}{\overline{\frac{\pi^2}{6}m_1^2}\cdots\overline{\frac{\pi^2}{6}m_s^2}},\end{aligned}$$

命 $q>0,a_1,\cdots,a_s$ 为整数, 则得

$$\frac{1}{q}\sum_{t=1}^{q}g\left(\frac{a_1t}{q},\cdots\frac{a_st}{q}\right)=\sum_{a_1m_1+\cdots+a_sm_s\equiv 0(\mathrm{mod}\ q)}\overset{\infty}{\underset{-\infty}{\cdots}}\sum\frac{1}{\overline{\frac{\pi^2}{6}m_1^2}\cdots\overline{\frac{\pi^2}{6}m_s^2}}.\tag{2.1}$$

主要的问题在于选取 $q,a_1,\cdots,a_s$ 使

$$H_1(q;a_1,\cdots,a_s)=\frac{1}{q}\sum_{t=1}^{q}g\left(\frac{a_1t}{q},\cdots,\frac{a_st}{q}\right)\tag{2.2}$$

尽可能的小.

最佳的途径很自然是这样的, 对于固定的 $q$, 在所有适合 $0\leqslant a_v<q$ 的整数中, 选取 $(a_1,\cdots,a_s)$ 使 $H_1(a_1,\cdots,a_s)$ 取极小. 无论如何, 这样做计算量太大而很难实现. 为此Коробов建议了下面的方法.

命 $q=p$ 为素数. 当 $z$ 经过 $1,2,\cdots,p-1$ 时, 在 $p-1$ 个和

$$H_1(p;1,z,\cdots,z^{s-1})$$

中, 寻求其最小者. 这一方法需要 $O(q^2)$ 次初等运算; 当 $q$ 较大时, 可以减低至 $O(q^{4/3})$

用代数数论的方法, 无需很复杂的计算, 即能给出一组 $q_1,a_1,\cdots,a_s$(用普通的台式计算机足够了). 然后再用 $O(q)$ 次初等运算, 即能得到 $H_1(q;a_1,\cdots,a_s)$ 之值.

命 $f(x_1,\cdots,x_s)$ 为每一变数都具有周期 1 的周期函数, 且有 Fourier 展开

$$f(x_1,\cdots,x_s)=\sum\overset{\infty}{\underset{-\infty}{\cdots}}\sum C(m_1,\cdots,m_s)e^{2\pi i(m_1x_1+\cdots+m_sx_s)},$$

此处

$$|C(m_1,\cdots,m_s)|<\frac{C}{(\bar{m}_1\cdots\bar{m}_s)^2},$$

这儿 $C$ 是一个正常数, 例如适合条件

$$\left|\frac{\partial^{2s}}{\partial x_1^2\cdots\partial x_s^2}f(x_1,\cdots,x_s)\right|<(2\pi)^{2s}C$$

的周期函数即属于这一函数类.

由于

$$\begin{aligned}\frac{1}{q}\sum_{t=1}^{q}f\left(\frac{a_1t}{q},\cdots,\frac{a_st}{q}\right)&=\sum\overset{\infty}{\underset{-\infty}{\cdots}}\sum C(m_1,\cdots,m_s)\frac{1}{q}\sum_{t=1}^{q}e^{2\pi i(a_1m_1+\cdots+a_sm_s)t/q}\\&=\sum_{a_1m_1+\cdots+a_sm_s\equiv0(\mathrm{mod}\ q)}\overset{\infty}{\underset{-\infty}{\cdots}}\sum C(m_1,\cdots,m_s),\end{aligned}$$

及

$$C(0,\cdots,0)=\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_s)\mathrm{d}x_1\cdots\mathrm{d}x_s,$$

所以

$$\begin{aligned}&\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_s)\mathrm{d}x_1\cdots\mathrm{d}x_s-\frac{1}{q}\sum_{t=1}^{q}f\left(\frac{a_1t}{q},\cdots,\frac{a_st}{q}\right)\\&\quad=-\sum_{a_1m_1+\cdots+a_sm_s\equiv0(\mathrm{mod}\ q)}\cdots\sum{}' C(m_1,\cdots,m_s),\end{aligned}\tag{2.3}$$

此处 $\Sigma\cdots\Sigma'$ 表示去掉 $m_1=\cdots=m_s=0$ 一项. 因此由 (2.1) 与 (2.2) 可知重积分

$$\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_s)\mathrm{d}x_1\cdots\mathrm{d}x_s$$

与单和

$$\frac{1}{q}\sum_{t=1}^{q} f\left(\frac{a_1 t}{q},\cdots,\frac{a_s t}{q}\right)$$

之差不超过

$$\begin{aligned}&\left|\sum_{a_1m_1+\cdots+a_sm_s\equiv 0(\bmod\ q)}\cdots\sum{}' C(m_1,\cdots,m_s)\right|\\ \leqslant & C\sum_{a_1m_1+\cdots+a_sm_s\equiv 0(\bmod\ q)}\cdots\sum{}' \frac{1}{(\bar{m}_1\cdots\bar{m}_s)^2}\\ \leqslant & C\left(\frac{\pi^2}{6}\right)^s [H_1(q;a_1,\cdots,a_s)-1].\end{aligned}$$

## §3. 全实代数数域的一些性质

命 $\mathscr{F}$ 表一 $n$ 次的全实代数数域, 其中一数 $\eta$ 的共轭数以 $\eta^{(1)}(=\eta),\eta^{(2)},\cdots,\eta^{(n)}$ 表之. 假定

$$\omega_1,\cdots,\omega_n$$

是其整底, 作方阵

$$\Omega=\begin{pmatrix}\omega_1^{(1)}, & \cdots, & \omega_n^{(1)}\\ & \cdots\cdots & \\ \omega_1^{(n)}, & \cdots, & \omega_n^{(n)}\end{pmatrix}\tag{3.1}$$

方阵

$$S=\Omega'\Omega=\left(\sum_{k=1}^{n}\omega_i^{(k)}\omega_j^{(k)}\right)\tag{3.2}$$

称为这个代数数域的基方阵. 显然基方阵是有有理整元素的方阵 (模群下, 基方阵的不变量是刻画代数数域的一个特征). 由 (3.2) 立得

$$\Omega^{-1}=S^{-1}\Omega'.\tag{3.3}$$

$S$ 的行列式即为域的基数.

如果 $\mathscr{F}$ 中有一单位 $\eta$, 其绝对值 $>1$, 而其共轭数适合于

$$|\eta^{(j)}|\leqslant c|\eta|^{-\frac{1}{n-1}},\tag{3.4}$$

这儿 $c$ 是一常数, 把 $\eta$ 表示为

$$\eta=h_1\omega_1+\cdots+h_n\omega_n,\tag{3.5}$$

由此及其共轭式子, 可以推出

$$(\eta^{(1)},\cdots,\eta^{(n)})=(h_1,\cdots,h_n)\Omega'. \tag{3.6}$$

因而

$$(\eta^{(1)},\cdots,\eta^{(n)})\Omega=(h_1,\cdots,h_n)S=(k_1,\cdots,k_n)\quad (\text{定义}).$$

即

$$\sum_{i=1}^{n}\eta^{(i)}\omega_j^{(i)}=k_j,$$

因而得出

$$|\eta\omega_j-k_j|\leqslant\sum_{i=2}^{n}|\eta^{(i)}|\,|\omega_j^{(i)}|\leqslant(n-1)c\omega|\eta|^{-\frac{1}{n-1}}, \tag{3.7}$$

这儿 $\omega=\max\limits_{i,j}|\omega_i^{(j)}|$.

命

$$1=a_1\omega_1+\cdots+a_n\omega_n. \tag{3.8}$$

则由 (3.7) 可以推得

$$|\eta-k|\leqslant A|\eta|^{-\frac{1}{n-1}}, \tag{3.9}$$

这儿

$$k=\sum_{i=1}^{n}a_ik_i,$$

常数 $A$ 仅与 $\mathscr{F}$ 及 $c$ 有关, 以后每次出现不一定是同一个 $A$.

因而得出 Diophantine 联立逼近式

$$\left|\omega_j-\frac{k_j}{k}\right|\leqslant\frac{A}{|k|^{n/(n-1)}}\quad(1\leqslant j\leqslant n). \tag{3.10}$$

这不是什么新结果. 我们的着眼点在于本节建议了一个实际算出 $k_i$ 的方法. 详细言之, 当 $n=2$ 时, 我们有连分数可用, 但 $n>2$ 的情况则迥然不同, 经典的方法只能证明存在无限多组 $k,k_1,\cdots,k_n$ 使 (3.10) 成立. 但一点也不能建议具体定出 $k,k_1,\cdots,k_n$ 的可行方法. 本节指出了寻求 $k,k_1,\cdots,k_n$ 的问题, 一变而为寻求全实域中适合于 (3.4) 的 $\eta$ 的问题.

如果知道了全实代数数域 $\mathscr{F}$ 的一个独立单位组, 则可以由以下的引理求出一个适合于 (3.4) 的 $\eta$ 的贯 $\{\eta_j\}$, 而且

$$|\eta_j|\to\infty\quad(\text{当 } j\to\infty).$$

因而我们有无穷多组 $k,k_1,\cdots,k_n$ 适合于 (3.10).

**引 3.1** 命

$$L_i(x_1,\cdots,x_m)=a_{i1}x_1+\cdots+a_{im}x_m \quad (1\leqslant i\leqslant m).$$

假定 $\det(a_{ij})\neq 0$, 则必有 $(x_1,\cdots,x_m)$ 使

$$L_i=L_j<K, \quad (1\leqslant i<j\leqslant m), \tag{3.11}$$

这儿 $K$ 是任给的实数. 又命

$$N=\max(|a_{11}|+\cdots+|a_{1m}|,\cdots,|a_{m1}|+\cdots+|a_{mm}|), \tag{3.12}$$

则必有整数组贯 $(y_1^{(t)},\cdots,y_m^{(t)})(t=1,2,\cdots)$ 使

$$L_i(y_1^{(t)},\cdots,y_m^{(t)})<-(2t-1)N \tag{3.13}$$

及

$$|L_i(y_1^{(t)},\cdots,y_m^{(t)})-L_j(y_1^{(t)},\cdots,y_m^{(t)})|\leqslant 2N. \tag{3.14}$$

**证** 解方程组

$$L_1=L_2=\cdots=L_m=K-1,$$

得出 $x_1,\cdots,x_m$, 代入 (3.11) 即合所需.

现在取 $K=-2Nt$ 及 $[x_i]=y_i^{(t)}(1\leqslant i\leqslant m)$, 则得

$$\begin{aligned}L_i(y_1^{(t)},\cdots,y_m^{(t)})&\leqslant L_i(x_1,\cdots,x_m)+\sum_{k=1}^{m}|a_{ik}|\\&<-2tN+N=-(2t-1)N.\end{aligned}$$

及

$$\begin{aligned}&|L_i(y_1^{(t)},\cdots,y_m^{(t)})-L_j(y_1^{(t)},\cdots,y_m^{(t)})|=\left|\sum_{k=1}^{m}(a_{ik}-a_{jk})y_k^{(t)}\right|\\&\leqslant\left|\sum_{k=1}^{n}(a_{ik}-a_{jk})x_k\right|+\sum_{k=1}^{n}|a_{ik}-a_{jk}|\\&=\sum_{k=1}^{n}|a_{ik}-a_{jk}|\leqslant 2N.\end{aligned}$$

引理证完.

**定理 3.1** 全实代数数域中存在单位贯 $\eta_t(=\eta_t^{(1)})(t=1,2,\cdots)$, 其共轭数 $\eta_t^{(2)},\cdots,\eta_t^{(n)}$ 都适合于

$$|\eta_i^{(i)}|<e^{-(2t-1)a} \quad (2\leqslant i\leqslant n) \tag{3.15}$$

及

$$e^{-2a}|\eta_t^{(j)}| \leqslant |\eta_t^{(i)}| \leqslant e^{2a}|\eta_t^{(j)}|, \quad (2 \leqslant i, j \leqslant n), \tag{3.16}$$

此处 $a$ 是一个仅与 $\mathscr{F}$ 有关的常数.

**证** 取 $\mathscr{F}$ 的一个独立单位组

$$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_r,$$

此处 $r = n - 1$.

命

$$\xi^{(i)} = \varepsilon_1^{(i)l_1} \cdots \varepsilon_r^{(i)l_r} \quad (2 \leqslant i \leqslant n).$$

取绝对值的对数得

$$\log|\xi^{(i)}| = \sum_{j=1}^{r} l_j \log|\varepsilon_j^{(i)}| \quad (2 \leqslant i \leqslant n).$$

命

$$a = \max_{2 \leqslant i \leqslant n} \left( \sum_{j=1}^{n} \log|\varepsilon_j^{(i)}| \right).$$

由于

$$\det(\log|\varepsilon_j^{(i)}|) \neq 0,$$

解方程

$$\log|\xi^{(2)}| = \cdots = \log|\xi^{(n)}| = -2at - 1,$$

得出 $l_1, \cdots, l_r$, 命

$$[l_i] = a_i^{(t)} \quad (1 \leqslant i \leqslant r).$$

则由引 3.1 可知单位

$$\eta_t = \varepsilon_1^{a_1^{(t)}} \cdots \varepsilon_r^{a_r^{(t)}},$$

合于定理所需.

## §4. 全实分圆域

命 $p \geqslant 5$ 表一素数, $n = \frac{1}{2}(p-1)$ 及 $r = n - 1 = \frac{1}{2}(p-3)$. 则全实分圆域 $R\left(2\cos\frac{2\pi}{p}\right)$ 是一个 $n$ 次域, 它的一组整底是

$$2\cos\frac{2\pi}{p}, \quad 2\cos\frac{4\pi}{p}, \cdots, 2\cos\frac{2\pi n}{p}. \tag{4.1}$$

悉知

$$\sum_{l=1}^{n} 2\cos\frac{2\pi l}{p} = -1, \tag{4.2}$$

及

$$\prod_{l=1}^{n} 2\cos\frac{2\pi l}{p} = (-1)^{[\frac{1}{2}(n+1)]} = (-1)^{\frac{p^2-1}{8}}. \tag{4.3}$$

命 $g$ 表示 $\bmod p$ 的一个原根. 由于

$$2\cos\frac{2\pi}{p}g^{l\pm n} = 2\cos\left(-\frac{2\pi}{p}g^l\right) = 2\cos\frac{2\pi}{p}g^l, \tag{4.4}$$

所以 (4.1) 可以写成

$$\omega_l = 2\cos\frac{2\pi}{p}g^{l+m} \quad (1 \leqslant l \leqslant n), \tag{4.5}$$

此处

$$|\omega_l| > |\omega_n| \quad (1 \leqslant l \leqslant n-1).$$

由 (4.4), 我们可以定义 $\omega_{l\pm n} = \omega_l$.

变换

$$\sigma : \omega_l \to \omega_{l+1}$$

是全实分圆域的一个自同构, 一共有 $n$ 个自同构

$$\sigma, \sigma^2, \cdots, \sigma^{n-1}, \sigma^n (= \sigma^0 = 1).$$

它们组成域的自同构群. 全实分圆域中一数 $\eta$, 经过自同构后, 变成它的 $n$ 个共轭数.

由于

$$\begin{aligned}\sum_{t=1}^{n} 2\cos\frac{2\pi lt}{p}\cdot 2\cos\frac{2\pi kt}{p} &= \sum_{t=1}^{n}\left(2\cos\frac{2\pi(l+k)t}{p} + 2\cos\frac{2\pi(l-k)t}{p}\right)\\ &= -2 + \begin{cases} p, & 若\ l = k;\\ 0, & 若\ l \neq k.\end{cases}\end{aligned}$$

因此

$$S = \begin{pmatrix} \omega_1, & \omega_2, & \cdots, & \omega_n\\ \omega_2, & \omega_3, & \cdots, & \omega_1\\ & \cdots\cdots & & \\ \omega_n, & \omega_1, & \cdots, & \omega_{n-1}\end{pmatrix}^2 = pI - 2M, \tag{4.6}$$

这儿 $M=(m_{ij}), m_{ij}=1$.

取 $R\left(2\cos\frac{2\pi}{p}\right)$ 的一个独立单位组 $\varepsilon_1,\cdots,\varepsilon_r$, 则由 §3 的方法, 取

$$\eta=\varepsilon_1^{l_1}\cdots\varepsilon_r^{l_r}, \tag{4.7}$$

使其 $n-1$ 个共轭数的绝对值都差不多大, 且都小于 1. 若

$$\eta=\sum_{j=1}^{n}h_j\omega_j, \tag{4.8}$$

则由 §3 可知

$$\left|\frac{k_i}{k}-\omega_i\right|=O\left(\frac{1}{|k|^{1+\frac{1}{n-1}}}\right)\quad(1\leqslant i\leqslant n), \tag{4.9}$$

此处

$$\begin{cases} k=-\sum_{j=1}^{n}h_j, \\ k_i=ph_i-2\sum_{j=1}^{n}h_j \end{cases}\quad(1\leqslant i\leqslant n). \tag{4.10}$$

已知 (见 Fricke[7])

$$\rho_l=\frac{\sin\frac{\pi}{p}g^{l+1}}{\sin\frac{\pi}{p}g^{l}}\quad(1\leqslant l\leqslant r) \tag{4.11}$$

是全实分圆域的一个独立单位组. 如果能用单位组

$$2\cos\frac{2\pi l}{p}\quad(1\leqslant l\leqslant r),$$

显然更为方便, 但它们并不总是独立的. 我们将在下节给出它们独立性的必要且充分的条件.

## §5. 全实分圆域的单位

**定理 5.1** 命 $\zeta_l=2\cos\frac{2\pi}{p}g^l$. 则 $\zeta_1,\cdots,\zeta_r$ 成为一组独立单位的必要且充分条件是

(i) 2 是 mod $p$ 的原根,

或

(ii) 2 的次数是 $n=\frac{1}{2}(p-1)$, 而且 $p\equiv 7 \pmod 8$.

**证**　若 $2 \equiv g^l \pmod p$, 则

$$\zeta_\mu = 2\cos\frac{2\pi}{p}g^\mu = 2\cos\frac{\pi}{p}g^{\mu+l}$$
$$= \frac{\sin\frac{\pi}{p}g^{\mu+2l}}{\sin\frac{\pi}{p}g^{\mu+l}} = \rho_{\mu+l}\cdots\rho_{\mu+2l-1}.$$

1) 若 2 为 mod $p$ 的原根, 则取 $g=2$, 从而 $l=1$, 所以

$$\zeta_\mu = \rho_{\mu+1}.$$

因此 $\zeta_1,\cdots,\zeta_r$ 是独立的.

2) 若 2 非 mod $p$ 的原根, 假定其次数为 $s$, 命 $p-1=ls$. 则取原根 $g$ 满足

$$g^l \equiv 2 \pmod p.$$

a) 若 $l|n$, 则命 $n=lt$, 所以

$$\zeta_1\zeta_{l+1}\cdots\zeta_{(t-1)l+1} = (\rho_{l+1}\cdots\rho_{2l})(\rho_{2l+1}\cdots\rho_{3l})\cdots(\rho_1\cdots\rho_l)$$
$$= \prod_{j=1}^{n}\rho_j = -1.$$

由于 $l>1$, 因此 $\zeta_1,\cdots,\zeta_r$ 是非独立的.

b) 若 $l|2n$, 而 $l\nmid n$, 则 $l=2m$. 命 $n=mt$.

当 $m\neq 1$ 时有

$$\zeta_1\zeta_{m+1}\cdots\zeta_{(t-1)m+1} = (\rho_{2m+1}\cdots\rho_{4m})(\rho_{3m+1}\cdots\rho_{5m})\cdots(\rho_{m+1}\cdots\rho_{3m})$$
$$= \prod_{j=1}^{n}\rho_j^2 = 1,$$

所以 $\zeta_1,\cdots,\zeta_r$ 是非独立的.

当 $m=1$ 时, 若有整数 $l_1,\cdots,l_n$ 使

$$\pm 1 = \zeta_1^{l_1}\cdots\zeta_n^{l_n}$$
$$= (\rho_3\rho_4)^{l_1}(\rho_4\rho_5)^{l_2}\cdots(\rho_n\rho_1)^{l_{n-2}}(\rho_1\rho_2)^{l_{n-1}}(\rho_2\rho_3)^{l_n}$$
$$= \rho_1^{l_{n-2}+l_{n-1}}\rho_2^{l_{n-1}+l_n}\rho_3^{l_n+l_1}\cdots\rho_n^{l_{n-3}+l_{n-2}},$$

则由 $\rho_1,\cdots,\rho_r$ 的独立性可知

$$l_{n-2}+l_{n-1} = l_{n-1}+l_n = l_n+l_1 = \cdots = l_{n-3}+l_{n-2}.$$

因此

$$l_1 = \cdots = l_n.$$

换言之, $\zeta_1, \cdots, \zeta_r$ 是独立的. 此时 2 的次数为 $n$, 且 $n$ 为奇数. 由

$$2^n \equiv \left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} (\bmod p)$$

可知 $p \equiv 7(\bmod 8)$.

定理证完.

## §6. 分点的选取与积分的数值误差

在全实分圆域 $R\left(2\cos\frac{2\pi}{p}\right)$ 中, 以 $\omega_1, \cdots, \omega_n$ 为整底, 其中 $|\omega_l| > |\omega_n| (1 \leqslant l \leqslant n-1)$. 又如 (4.10) 来定义 $k, k_1, \cdots, k_n$.

定义

$$q = |k|, a_1 = 1, a_{j+1} = |k_j| \quad (1 \leqslant j \leqslant r). \tag{6.1}$$

我们就用下面的渐近公式来逼近重积分:

$$\int_0^1 \cdots \int_0^1 f(x_1, \cdots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}n_x \approx \frac{1}{q} \sum_{t=1}^{q} f\left(\frac{a_1 t}{q}, \cdots \frac{a_n t}{q}\right). \tag{6.2}$$

总结一下: 取 $R\left(2\cos\frac{2\pi}{p}\right)$ 的一组独立单位 $\varepsilon_1, \cdots, \varepsilon_r$. 并取

$$\eta = \varepsilon_1^{l_1} \cdots \varepsilon_r^{l_r},$$

使得其 $r$ 个共轭数的绝对值都差不多大, 且都小于 1. 把 $\eta$ 用整底表出

$$\eta = h_1\omega_1 + \cdots + h_n\omega_n. \tag{6.3}$$

则

$$q = \left|\sum_{j=1}^{n} h_j\right|, \quad a_1, = 1, \quad a_{j+1} = \left|p h_j - 2\sum_{j=1}^{n} h_j\right| \quad (1 \leqslant j \leqslant r) \tag{6.4}$$

就给出了我们的分点.

现在举出以下的数值例子.

**例 1** 取 $R\left(2\cos\frac{2\pi}{5}\right)$ 及 $\varepsilon_1 = 2\cos\frac{4\pi}{5}, \varepsilon_2 = 2\cos\frac{2\pi}{5}$. 则由

$$\varepsilon_1^n = q_0^{(n)}\varepsilon_2 - q_1^{(n)},$$

即得

$$H_1(|q_0^{(n)}|;1,|q_1^{(n)}|)=1+O\left(\frac{\log|q_0^{(n)}|}{|q_0^{(n)}|^2}\right)^{[2]}.$$

**例 2**　取 $R\left(2\cos\frac{2\pi}{7}\right)$ 及

$$\varepsilon_1=2\cos\frac{6\pi}{7},\quad \varepsilon_2=2\cos\frac{2\pi}{7},\quad \varepsilon_3=2\cos\frac{4\pi}{7}.$$

由方程

$$|\varepsilon_2^{\alpha}\varepsilon_3^{\beta}|=|\varepsilon_3^{\alpha}\varepsilon_1^{\beta}|$$

解得

$$\frac{\alpha}{\beta}=1.356\cdots\doteq\frac{4}{3}.$$

由展开式

$$\varepsilon_1^8\varepsilon_2^6=-227\varepsilon_1-45\varepsilon_2-146\varepsilon_3$$

得出

$$q=418,\quad a_1=1,\quad a_2=335,\quad a_3=103.$$

由实际计算得

$$H_1(418;1,335,103)\leqslant 1.0108146.$$

**例 3**　取 $R\left(2\cos\frac{2\pi}{11}\right)$ 及

$$\varepsilon_1=2\cos\frac{10\pi}{11},\quad \varepsilon_2=2\cos\frac{2\pi}{11},\quad \varepsilon_3=2\cos\frac{8\pi}{11},\quad \varepsilon_4=2\cos\frac{4\pi}{11},\quad \varepsilon_5=2\cos\frac{6\pi}{11}.$$

解方程组

$$|\varepsilon_2^{\alpha}\varepsilon_4^{\beta}\varepsilon_5^{\gamma}\varepsilon_3^{\delta}|=|\varepsilon_3^{\alpha}\varepsilon_5^{\beta}\varepsilon_2^{\gamma}\varepsilon_1^{\delta}|=|\varepsilon_4^{\alpha}\varepsilon_3^{\beta}\varepsilon_1^{\gamma}\varepsilon_5^{\delta}|=|\varepsilon_5^{\alpha}\varepsilon_1^{\beta}\varepsilon_4^{\gamma}\varepsilon_2^{\delta}|$$

得出

$$\frac{\alpha}{\beta}\doteq 1.412\doteq\frac{7}{5},\quad \frac{\beta}{\delta}\doteq 1.584\doteq\frac{8}{5},\quad \frac{\gamma}{\delta}\doteq 0.944\doteq\frac{5}{5}.$$

由展开式

$$\varepsilon_1^7\varepsilon_2^8\varepsilon_3^5\varepsilon_4^5=-3345\varepsilon_1-271\varepsilon_2-2825\varepsilon_3-998\varepsilon_4-1950\varepsilon_5$$

得出

$$q=9389,\quad a_1=1,\quad a_2=8628,\quad a_3=6408,$$
$$a_4=2908,\quad a_5=7800.$$

经过计算得

$$H_1(q;a_1,\cdots,a_5)\leqslant 1.0081175.$$

**例 4** 取 $R\left(2\cos\frac{2\pi}{13}\right)$ 及

$$\varepsilon_1=2\cos\frac{12\pi}{13},\quad \varepsilon_2=2\cos\frac{2\pi}{13},\quad \varepsilon_3=2\cos\frac{10\pi}{13},\quad \varepsilon_4=2\cos\frac{4\pi}{13},$$
$$\varepsilon_5=2\cos\frac{8\pi}{13},\quad \varepsilon_6=2\cos\frac{6\pi}{13}.$$

由展开式

$$\varepsilon_1^7\varepsilon_2^8\varepsilon_3^5\varepsilon_4^6\varepsilon_5^4=-12494\varepsilon_1-726\varepsilon_2-11084\varepsilon_3-2738\varepsilon_4$$
$$-8587\varepsilon_5-5575\varepsilon_6$$

得出

$$q=41204,\quad a_1=1,\quad a_2=38810,\quad a_3=31766\quad a_4=20480,$$
$$a_5=5610,\quad a_6=29223.$$

经计算得

$$H_1(q;a_1,\cdots,a_6)\leqslant 1.0094250.$$

**例 5** 取 $R\left(2\cos\frac{2\pi}{23}\right)$ 及

$$\varepsilon_1=2\cos\frac{22\pi}{23},\quad \varepsilon_2=2\cos\frac{2\pi}{23},\quad \varepsilon_3=2\cos\frac{20\pi}{23},\quad \varepsilon_4=2\cos\frac{4\pi}{23},$$
$$\varepsilon_5=2\cos\frac{18\pi}{23},\quad \varepsilon_6=2\cos\frac{6\pi}{23},\quad \varepsilon_7=2\cos\frac{16\pi}{23},\quad \varepsilon_8=2\cos\frac{8\pi}{23},$$
$$\varepsilon_9=2\cos\frac{14\pi}{23},\quad \varepsilon_{10}=2\cos\frac{10\pi}{23},\quad \varepsilon_{11}=2\cos\frac{12\pi}{23}.$$

由展开式

$$\varepsilon_1^5\varepsilon_2^6\varepsilon_3^4\varepsilon_4^7\varepsilon_5^4\varepsilon_6^3\varepsilon_7^4\varepsilon_8^4\varepsilon_9^3\varepsilon_{10}^2=-120834\varepsilon_1-2251\varepsilon_2-116374\varepsilon_3$$
$$-8837\varepsilon_4-107785\varepsilon_5-19269\varepsilon_6-95704\varepsilon_7-32774\varepsilon_8$$
$$-81027\varepsilon_9-48350\varepsilon_{10}-64842\varepsilon_{11}$$

得出

$$q=698047,\quad a_1=1,\quad a_2=685041,\quad a_3=646274,$$
$$a_4=582461,\quad a_5=494796,\quad a_6=384914,\quad a_7=254860,$$
$$a_8=107051,\quad a_9=642292,\quad a_{10}=467527,\quad a_{11}=284044.$$

经计算得

$$H_1(q;a_1,\cdots,a_{11})\leqslant 1.2333543.$$

**附记** 其他的全实域亦可以用来定出计算多重积分的分点. 例如 Dirichlet 域, 即若干次二次扩张所成的域 (见华罗庚与王元 [3,4]). 但我们猜测分圆域不但易于计算, 而且在同次数的域中, 它能给出最好的结果.

例如用 Dirichlet 域 $R(\sqrt{2}, \sqrt{5})$. 取

$$\varepsilon_1 = \frac{1+\sqrt{5}}{2}, \quad \varepsilon_2 = 1+\sqrt{2}, \quad \varepsilon_3 = 3+\sqrt{10}.$$

展开

$$\begin{aligned}\varepsilon_1^6\varepsilon_2^4\varepsilon_3^2 &= (9+4\sqrt{5})(17+12\sqrt{2})(19+6\sqrt{10}) \\ &= 5787+4092\sqrt{2}+2588\sqrt{5}+1830\sqrt{10},\end{aligned}$$

得出

$$H_1(5787; 1, 2397, 1366, 939) \leqslant 1.001498.$$

## 参 考 文 献

[1] 华罗庚, 堆垒素数论, 科学出版社, 1957.

[2] 华罗庚与王元, 数值积分及其应用, 科学出版社, 1963.

[3] Hua Loo Keng and Wang Yuan, On Diophantine approximations and numerical integrations (I), *Sci, Sinica*, **13**, 6, 1964, 1007~1009.

[4] Hua Loo Keng and Wang Yuan, On Diophantine approximations and numerical integrations (II), *Sci. Sinica*, **13**, 6, 1964, 1009~1010.

[5] Коробов, Н. М., Теоретикочисловые методы в приближенном анализе, ГИ ФМЛ, 1963.

[6] Halton, J. H., On the efficiency of certain quasirandom sequences of points in evaluating multidimensional integrals, *Num. Math.* **27**, 2(1960), 84~90.

[7] Fricke, R., Lehrbuch der Algebra III, 1928.

# 关于一类函数的插入公式 *

王 元

(中国科学院数学研究所)

I. 命 $E_s^\alpha(C)$ 表示适合下面条件的函数构成的函数类:

$$f(x_1,\cdots,x_s)=\sum\mathop{\cdots}\limits_{-\infty}^{\infty}\sum\times C(m_1,\cdots,m_s)e^{2\pi i(m_1x_1+\cdots+m_sx_s)},$$

其中

$$|C(m_1,\cdots,m_s)|\leqslant\frac{C}{(\bar{m}_1\cdots\bar{m}_s)^\alpha},$$

此处 $\bar{m}=\max(1,|m|)$, 而 $\alpha>1$ 及 $C>0$ 均为绝对常数. 若 $f\in E_s^\alpha(C)$, 则定义 $v$ 如下:

$$v(x_1,\cdots,x_s)=\sum\mathop{\cdots}\limits_{-\infty}^{\infty}\sum B(m_1,\cdots,m_s)C(m_1,\cdots,m_s)e^{2\pi i(m_1x_1+\cdots+m_sx_s)},$$

此处 $C(m_1,\cdots,m_s)$ 为 $f$ 的 Fourier 系数, 而 $B(m_1,\cdots,m_s)$ 适合

$$|B(m_1,\cdots,m_s)|\leqslant\frac{1}{(\bar{m}_1\cdots\bar{m}_s)^w}\quad(0\leqslant\omega\leqslant1).$$

命 $N$ 为素数及

$$N_1=[N^{\frac{\alpha}{2\alpha-1}}(\ln N)^{\frac{-(s-1)(\alpha-1)}{2\alpha-1}}]+1.$$

又命

$$\tilde{C}(m_1,\cdots,m_s)=\frac{1}{N}\sum_{k=1}^{N}f\left(\frac{a_1k}{N},\cdots,\frac{a_sk}{N}\right)e^{-2\pi i\frac{a_1m_1+\cdots+a_sm_s}{N}k},$$

$$Q(x_1,\cdots,x_s)=\sum_{\bar{m}_1\cdots\bar{m}_s\leqslant N_1}B(m_1,\cdots,m_s)\times\tilde{C}(m_1,\cdots,m_s)e^{2\pi i(m_1x_1+\cdots+m_sx_s)},$$

及

$$\varDelta=\min_a\sup_{f\in E_s^\alpha(C)}|v(x_1,\cdots,x_s)-Q(x_1,\cdots,x_s)|,$$

* 原载:《科学通报》, **9**, 1966, 389~391.

此处 $a=(a_1,\cdots,a_s)$ 为以整数为分量的矢量.

本文的目的在于证明次之二定理.

**定理 1** 下面的估计式成立

$$\Delta \leqslant c_1(\alpha,s)CN^{-\frac{\alpha(\alpha+\omega-1)}{2\alpha-1}}(\ln N)^{\frac{(s-1)(\alpha^2+\alpha\omega-\omega)}{2\alpha-1}},$$

此处 $c_1(\alpha,s)$ 为依赖于 $\alpha$ 及 $s$ 的常数.

命 $f\in E_s^\alpha(C)$, 且对每一变数 $x_1,\cdots,x_s$ 都是奇函数. 又命 $u(x_1,\cdots,x_s)$ 在 $s$ 维单位立方体 $G_s$ 中适合 Poisson 方程, 而在 $G_s$ 的边界上取零值. 则由定理 1 可以推出[1,2]:

**定理 2** 命 $s\geqslant 2$. 则对于素数 $N$, 皆存在整数矢量 $a=(a_1,\cdots,a_s)$ 使

$$\left|u(x_1,\cdots,x_s)-\sum_{k=1}^{N} f\left(\frac{a_1k}{N},\cdots,\frac{a_k}{N}\right)\psi_k(x_1,\cdots,x_s)\right|$$
$$\leqslant c_2(\alpha,s)N^{\frac{-\alpha(\alpha s+2-s)}{(2\alpha-1)s}}(\ln N)^{\frac{(s-1)(\alpha^2 s+2\alpha-2)}{(2\alpha-1)s}},$$

此处

$$\psi_k(x_1,\cdots,x_s)=-\frac{1}{4\pi N}$$
$$\times\sum_{\bar m_1\cdots\bar m_s\leqslant N_1}\frac{e^{2\pi i\left(m_1\left(x_1-\frac{a_1k}{N}+\cdots+m_s\left(x_s-\frac{a_sk}{N}\right)\right)\right)}}{m_1^2+\cdots+m_s^2},$$

其中 $\Sigma'$ 表示去掉 $m_1=\cdots=m_s=0$ 一项.

定理 1 与定理 2 改进了以往Коробов[1,2] 与作者[3] 的结果. 例如当 $\alpha=s=2$ 时, 由定理 2 得出误差 $O(N^{-\frac{4}{3}+\varepsilon})$, 而由Коробов及作者原来的结果只能分别得出 $O(N^{-1+\varepsilon})$ 及 $O(N^{-\frac{6}{5}+\varepsilon})$.

II. 定理 1 的证明需要次之二引理.

**引理 1**[4] 命 $l_i(1\leqslant i\leqslant s)$ 及 $n_1$ 为满足 $\bar l_1\cdots\bar l_s\geqslant 3^s$ 及 $1\leqslant n_1\leqslant \bar l_1\cdots\bar l_s/3^s$ 的整数, 而 $\alpha>1$ 为实数. 则

$$\sum_{\bar m_1\cdots\bar m_s\leqslant n_1}\frac{1}{((\overline{l_1+m_1})\cdots(\overline{l_s+m_s}))^\alpha}\leqslant c_3(\alpha,s)\frac{n_1^\alpha}{(\bar l_1\cdots\bar l_s)^\alpha}.$$

**引理 2**[5] 命 $\alpha>1$ 及 $N$ 为素数, 则存在整数矢量 $a=(a_1,\cdots,a_s)$ 使

$$\sum_{a_1m_1+\cdots+a_sm_s\equiv 0(\mathrm{mod}\ N)}{}' \frac{1}{(\bar m_1\cdots\bar m_s)^\alpha}\leqslant c_4(\alpha,s)\frac{(\ln N)^{\alpha(s-1)}}{N^\alpha}.$$

**定理 1 的证明:**

1) $$|v-Q|\leqslant\sum_{\bar m_1\cdots\bar m_s\leqslant N_1}|B(m_1,\cdots,m_s)||C(m_1,\cdots,m_s)-\tilde C(m_1,\cdots,m_s)|$$
$$+\sum_{\bar m_1\cdots\bar m_s>N_1}|B(m_1,\cdots,m_s)||C(m_1,\cdots,m_s)|=\sum\nolimits_1+\sum\nolimits_2.$$

2) 由于

$$C(m_1,\cdots,m_s)-\tilde C(m_1,\cdots,m_s)$$
$$=-\sum_{a_1l_1+\cdots+a_sl_s\equiv0(\mathrm{mod}\ N)}{}' C(l_1+m_1,\cdots,l_s+m_s),$$

所以

$$\left|\sum\nolimits_1\right|\leqslant\sum_{\bar m_1\cdots\bar m_s\leqslant N_1}\frac{1}{(\bar m_1\cdots\bar m_s)^\omega}$$
$$\times\sum_{a_1l_1+\cdots+a_sl_s\equiv0(\mathrm{mod}\ N)}{}'\frac{C}{(\overline{l_1+m_1})\cdots(\overline{l_s+m_s})^\alpha}.$$

由引理 2 可知, 当 $N>c_5(\alpha,s)$ 时, 同余式

$$a_1l_1+\cdots+a_sl_s\equiv0(\mathrm{mod}\ N)$$

的非零解适合

$$\bar l_1\cdots\bar l_s\geqslant c_6(\alpha,s)\frac{N}{(\ln N)^{s-1}}>\frac{N_1}{3^s}$$

(当 $N\leqslant c_5(\alpha,s)$ 时, 定理显然成立). 所以由引理 1 与引理 2 得

$$\left|\sum\nolimits_1\right|\leqslant C\sum_{k=0}^{[\log_2N_1]}\sum_{2^{-k-1}N_1<\bar m_1\cdots\bar m_s\leqslant2^{-k}N_1}\frac{1}{(\bar m_1\cdots\bar m_s)^\omega}$$
$$\times\sum_{a_1l_1+\cdots+a_sl_s\equiv0(\mathrm{mod}\ N)}{}'\frac{1}{((\overline{l_1+m_1})\cdots(\overline{l_s+m_s}))^\alpha}$$
$$\leqslant C\sum_{k=0}^{[\log_1N_1]}(2^{-k-1}N_1)^{-\omega}\sum_{a_1l_1+\cdots+a_sl_s\equiv0(\mathrm{mod}\ N)}{}'\sum_{\bar m_1\cdots\bar m_s\leqslant2^{-k}N_1}$$
$$\frac{1}{(\overline{l_1+m_1})\cdots(\overline{l_s+m_s})^\alpha}$$
$$\leqslant c_3(\alpha,s)C\sum_{k=0}^{[\log_2N_1]}(2^{-k-1}N_1)^{-\omega}\sum_{a_1l_1+\cdots+a_sl_s\equiv0(\mathrm{mod}\ N)}{}'\frac{(2^{-k}N_1)^\alpha}{(\bar l_1\cdots\bar l_s)^\alpha}$$
$$\leqslant c_7(\alpha,s)CN_1^{-\omega+\alpha}N^{-\alpha}(\ln N)^{\alpha(s-1)}\sum_{k=0}^{\infty}(2^{\alpha-1})^{-k}$$

$$\leqslant c_8(\alpha,s)CN^{\frac{-\alpha(\alpha+\omega-1)}{2\alpha-1}}(\ln N)^{\frac{(s-1)(\alpha^{+}\alpha\omega-\omega)}{2\alpha-1}}.$$

3) $$\left|\sum\nolimits_2\right| \leqslant \sum_{\bar{m}_1\cdots\bar{m}_s>N_1}\frac{C}{(\bar{m}_1\cdots\bar{m}_s)^{\alpha+\omega}} \leqslant c_9(\alpha,s)CN_1^{-\alpha-\omega+1}(\ln N_1)^{s-1}$$

$$\leqslant c_{10}(\alpha,s)CN^{\frac{-\alpha(\alpha+\omega-1)}{2\alpha-1}}(\ln N)^{\frac{(s-1)(\alpha^2+\alpha\omega-\omega)}{2\alpha-1}}.$$

由 1)、2)、3) 即得定理.

## 参 考 文 献

[1] Коробов Н. М., Вопросы вычисл. матем. ц вычисл. техн., Машгиз, 1963.

[2] Коробов Н. М., Теоретикочисловые методы в приближенном анализе, Гифмл, 1963.

[3] Wang Yuan, *Sci. Sinica*, **14**, 4, 629~631(1965).

[4] Wang Yuan, *Sci. Sinica*, **10**, 6, 632~636(1961).

[5] Бахвалов Н. С., Вестник МГУ, [4], 3~18(1959).

# 论一致分布与近似分析 —— 数论方法 (I)*

华罗庚　王　元

(中国科学院数学研究所)

## 摘　要

本文研究由实分圆域定义的一致分布点列贯, 求出了它们的偏差, 并应用于数值积分问题.

## §1. 序　言

命 $G_s$ 是 $s$ 维空间的单位立方体.

$$0 \leqslant x_1 \leqslant 1, \cdots, 0 \leqslant x_s \leqslant 1.$$

令 $n_1 < n_2 < \cdots$ 为正整数贯及

$$P_{n_l}(j) = (x_1^{n_l}(j), \cdots, x_s^{n_l}(j)) \quad (1 \leqslant j \leqslant n_l)$$

表示 $G_s$ 中的点集. 对于任何 $(\gamma_1, \cdots, \gamma_s) \in G_s$, 命 $N_{n_l}(\gamma, \cdots, \gamma_s)$ 表示点列 $P_{n_l}(j)(1 \leqslant j \leqslant n_l)$ 中适合不等式

$$0 \leqslant x_s^{(n_l)}(j) < \gamma_1, \cdots, 0 \leqslant x_s^{(n_l)}(j) < \gamma_s$$

的个数. 若

$$\lim_{l \to \infty} \frac{N_{n_l}(\gamma_1, \cdots, \gamma_s)}{n_l} = \gamma_1, \cdots, \gamma_s,$$

则称点集贯 $P_{n_l}(j)(n_1 < n_2 < \cdots)$ 在 $G_s$ 上一致分布 [1]. 进而言之, 若满足更强的条件:

$$\left| \frac{N_{n_l}(\gamma_1, \cdots, \gamma_s)}{n_l} - \gamma \cdots \gamma_s \right| < \varphi(n_l),$$

此处 $\varphi(n_l) = o(1)$, 则称点集贯 $P_{n_l}(j)(n_1 < n_2 < \cdots)$ 一致分布, 且有偏差 $\varphi(n)$. 特别当 $n_l = l$ 及 $x_1^{(l)}(j) = x_1(j), \cdots, x_s^{(l)}(j) = x_s(j)(l = 1, 2, \cdots)$ 时, 则称贯 $P(j) = (x_1(j), \cdots, x_s(j))(j = 1, 2, \cdots)$ 在 $G_s$ 上一致分布.

* 原载《中国科学》, 16,4,1973,339～357.

命 $p \geqslant 5$ 为一素数, $r = \frac{1}{2}(p-1)$ 及 $q = r - 1 = \frac{1}{2}(p-3)$. 实分圆域 $\mathscr{R}_r = R\left(2\cos\frac{2\pi}{p}\right)$ 为一个 $r$ 次的代数数域. 命

$$\omega_1 = 2\cos\frac{2\pi}{p}, \omega_2 = 2\cos\frac{4\pi}{p}, \cdots, \omega_q = \cos\frac{2\pi q}{p}.$$

又命

$$\left|\frac{h_i^{(l)}}{n_l} - \omega_i\right| \leqslant c(\mathscr{R}_r) n_l^{-1-\frac{1}{q}} (1 \leqslant i \leqslant q)$$

表示诸 $\omega_i$ 的联立有理逼近, 此处我们用 $c(f, \cdots, g)$ 表示仅与 $f, \cdots, g$ 有关的正常数, 但不一定取相同的数值.

**定理 1**　命

$$P(j) = (\{\omega_1 j\}, \cdots, \{\omega_q j\})(j = 1, 2, \cdots), \tag{1.1}$$

其中 $\{x\}$ 表示 $x$ 的分数部分. 则贯 $P(j)(j = 1, 2, \cdots)$ 有偏差

$$\varphi(n) = c(\mathscr{R}_r, \varepsilon) n^{-1+\varepsilon},$$

此处 $\varepsilon$ 为任给正数, 本文将沿用这一规定.

**定理 2**　命

$$P_{n_l}(j) = \left(\left\{\frac{j}{n_l}\right\}, \left\{\frac{h_1^{(l)} j}{n_l}\right\}, \cdots, \left\{\frac{h_q^{(l)} j}{n_l}\right\}\right) \quad (1 \leqslant j \leqslant n_l). \tag{1.2}$$

则点集贯 $(P_{n_l}(j))$ 有偏差

$$\varphi(n) = c(\mathscr{R}_r, \varepsilon) n^{-\frac{1}{2} - \frac{1}{2q} + \varepsilon}.$$

命 $E_s^\alpha(C)$ 表示函数类

$$f(x_1, \cdots, x_s) = \sum \overset{\infty}{\underset{-\infty}{\cdots}} \sum C(m_1, \cdots, m_s) e^{2\pi i(m_1 x_1 + \cdots + m_s x_s)},$$

此处 Fourier 系数满足

$$|C(m_1, \cdots, m_s)| \leqslant \frac{C}{(\bar{m}_1 \cdots \bar{m}_s)^\alpha},$$

其中 $\alpha > 0$ 与 $C > 0$ 均为绝对常数及 $\bar{m} = \max(1, |m|)$.

在本文中我们假定 $\alpha > 1$.

将点集贯 (1.1) 与 (1.2) 式用于数值积分问题, 我们得下面结果:

**定理 3** 命 $l$ 为 $\geqslant \alpha$ 的最小整数及 $\mu_{n,l,i}$ 为由下式定义的整数集

$$\left(\sum_{j=-n}^{n} z^j\right)^l = \sum_{j=-nl}^{nl} \mu_{n,l,iz^j}.$$

则得

$$\sup_{f\in E_q^\alpha(C)}\left|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_q)dx_1\cdots dx_q - \frac{1}{(2n+1)^l}\right.$$
$$\left.\times\sum_{j=-nl}^{nl}\mu_{n,l,j}f(\omega_1 j,\cdots,\omega_q j)\right| \leqslant C\cdot c(\mathscr{R}_r,\alpha,\varepsilon)n^{-\alpha+\varepsilon}. \qquad (1.3)$$

**定理 4** 我们有

$$\sup_{f\in E_r^\alpha(C)}\left|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_r)\mathrm{d}x_1\cdots\mathrm{d}x_r - \frac{1}{n}\sum_{j=1}^{n_l} f\left(\frac{j}{n},\frac{h_l^{(l)}j}{n_l},\cdots,\frac{h_q^{(l)}j}{n_l}\right)\right|$$
$$\leqslant C\cdot c(\mathscr{R}_r,\varepsilon)n_l^{-\frac{\alpha}{2}-\frac{\alpha}{2q}+\varepsilon}. \qquad (1.4)$$

悉知在某些条件之下, 非周期函数的积分可以化为周期函数的积分来计算 (见文献 [2]). 当 $\alpha=2$ 时, 我们还可以用下面的简单公式来代替 (1.3) 式.

$$\sup_{f\in E_q^2(C)}\left|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_r)\mathrm{d}x_1\cdots\mathrm{d}x_r - \frac{1}{n}\sum_{j=-n}^{n_l}\left(1-\frac{|j|}{n}\right)f(\omega_1 j,\cdots,\omega_q j)\right|$$
$$\leqslant C\cdot c(\mathscr{R}_r,\varepsilon)n^{-2+\varepsilon}. \qquad (1.5)$$

由 Roth[3] 关于一致分布点集贯的偏差的 $\Omega$ 结果, 即可由定理 1 给出估计, 除因子 $n^\varepsilon$ 之外, 已不允许作本质的改进了. 我们称这种点集贯的"最佳分布"的点集贯. 其他著名的"最佳分布"点集贯在 1959 年与 1960 年分别由Коробов[2] 与 Halton[4] 引入. 我们也可以证明由定理 3 给出的误差项是不能作本质的改进的 (见文献 [5]). 欲要得到 (1.2) 式的整数贯 $(h_1^{(1)},\cdots,h_q^{(l)};n_l)$, 只要进行 $c(\mathscr{R}_r)logn_l$ 次初等运算.

古典的数值积分方法是借助于点集贯

$$\left(\frac{j_1}{m},\cdots,\frac{j_s}{m}\right) \quad (0\leqslant j_1,\cdots,j_s\leqslant m-1) \qquad (1.6)$$

来建立的. 即在 $G_s$ 上的积分用和

$$\frac{1}{m^s}\sum_{j_1,\cdots,j_s} f\left(\frac{j_1}{m},\cdots,\frac{j_s}{m}\right) \qquad (1.7)$$

来近似计算. (1.6) 式的点数为 $n = m^s$. 易证 (1.6) 的偏差 $\geqslant n^{-\frac{1}{s}}$ 及 $E_s^a(C)$ 上函数在 $G_s$ 上, 用古典求积公式所得的误差项 $\geqslant 2C_n^{-a/s}$(取 $f(x_1,\cdots,x_s) = C(e^{2\pi i m_1 x_1} + e^{-2\pi i m_1 x_1})/m^a$), 则

$$\int_0^1 \cdots \int_0^1 f(x_1,\cdots,x_s)\mathrm{d}x_1\cdots\mathrm{d}x_s = 0,$$

及

$$\frac{1}{n}\sum_{j_1,\cdots,j_s} f\left(\frac{j_1}{m},\cdots,\frac{j_s}{m}\right) = 2Cn^{-a/s)}.$$

Бахвалов[6] 与作者 [7] 借助于 Fibonacci 贯 $(F_l)$, 对 $q=1$, 独立地证明了公式 (1.4) 式, 并得到误差项为 $O(F_l^{-a}\log 3F_l)$, 此处

$$F_l = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{l+1} - \left(\frac{1-\sqrt{5}}{2}\right)l+1\right]\ (l \geqslant 0).$$

因 $\mathscr{R}_r = R\left(2\cos\frac{2\pi}{5}\right) = R(\sqrt{5})$, 所以点集贯 (1.1) 与 (1.2) 式分别可以看作由黄金分割与 Fibonacci 贯得来的点集贯

$$\left\{\frac{\sqrt{5}-1}{2}j\right\}(j=1,2,\cdots)\text{与}\left(\left\{\frac{j}{F_l}\right\},\left\{\frac{F_{l-1}j}{F_1}\right\}\right)(1\leqslant j\leqslant F_l) \tag{1.8}$$

的推广. 因此我们曾建议用实分圆域来处理高维空间的数值积分问题, 并给出了 (1.4) 式对于 $2 \leqslant q \leqslant 10$ 时的若干数值例子 [7−9]. 虽然我们可用其他完实域来代替分圆域, 例如用 Dirichlet 域 $R(\sqrt{p_1},\cdots,\sqrt{p_h})$, 此处诸 $p_i$ 为 $h$ 个互不相同的素数, 但我们猜想分圆域在同次域中常能给出最佳的结果, 而且也便于计算.

上述诸定理的证明依赖于 W.M.Schmidt[10] 关于代数数的联立有理逼近的重要结果.

**引理 1.1**　命 $\alpha_1,\cdots,\alpha_s$ 为实代数数及 $1,\alpha,\cdots,\alpha_s$ 在有理数域 $R$ 上线性独立. 则

$$\langle \alpha_1 m_1 + \cdots + \alpha_s m_s\rangle \geqslant c(\alpha_1,\cdots,\alpha_s,\varepsilon)(\bar{m}_1\cdots\bar{m}_s)^{-1-\varepsilon},$$

其中 $m_1,\cdots,m_s$ 是一组不全为零的整数及 $\langle x\rangle = \min(\{x\}, 1-\{x\})$.

我们也引用了由Бахвалов[6], Коробов[2], Haselgrove[11] 与 Hlawka[12,13] 等人引入的近似分析中的数论方法.

用 A. Baker[14] 的类似结果来代替引理 1.1, 我们也可以得到相应的近似分析结果. 例如可以证明, 贯

$$P(j) = (\{ej\},\{e^2 j\},\cdots,\{e^s j\})\quad (j=1,2,\cdots) \tag{1.9}$$

有偏差 $\varphi(n)=c(s,\varepsilon)n^{-1+\varepsilon}$.

用 Wang 520 台式电子计算机, 我们得到 (1.5) 式的例子

$$\sup_{f\in E_4^2(C)}\left|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_4)dx_1\cdots dx_4-\frac{1}{3000}\sum_{j=-3000}^{3000}\left(1-\frac{|j|}{3000}\right)\right.$$
$$\left.\times f\left(2j\cos\frac{2\pi}{11},2j\cos\frac{4\pi}{11},2j\cos\frac{6\pi}{11},2j\cos\frac{8\pi}{11}\right)\right|\leqslant 0.065C.$$

## §2. 完实代数数域

命 $\mathscr{F}_s$ 表示一个 $s$ 次的完实代数数域, 对于 $\mathscr{F}_s$ 中的一个数 $\eta$, 命 $\eta^{(1)}(=\eta),\eta^{(2)},\cdots,\eta^{(s)}$ 表示其共轭. 假定 $\omega_1,\cdots,\omega_s$ 是 $\mathscr{F}_s$ 的整底, 作方阵

$$\Omega=(\omega_j^{(i)})(1\leqslant i,j\leqslant s)$$

则方阵

$$S=\Omega'\Omega=\left(\sum_{k=1}^s\omega_i^{(k)}\omega_j^{(k)}\right)(1\leqslant i,j\leqslant s)$$

称为 $\mathscr{F}_s$ 的基方阵. 显然基方阵是有有理整元素的方阵. 在模群作用下, 基方阵的不变性是表示代数数域的一个特征. $S$ 的行列式 $\det S$ 称为域的基数.

若 $\mathscr{F}_s$ 中有一个单位 $\eta$, 适合

$$|\eta|>1,|\eta^{(j)}|\leqslant c(\mathscr{F}_s,)|\eta|^{-\frac{1}{s-1}}\quad(2\leqslant j\leqslant s),\tag{2.1}$$

则将 $\eta$ 表为

$$\eta=\sum_{i=1}^s k_l\omega_i,\tag{2.2}$$

此处诸 $k_i$ 为有理整数. 由 (2.2) 式及其共轭式子可以推出.

$$(\eta^{(1)},\cdots,\eta^{(s)})=(k_1,\cdots,k_s)\Omega'.\tag{2.3}$$

因而定义

$$(\eta^{(1)},\cdots,\eta^{(s)})\Omega=(k_1,\cdots,k_s)S=(h_1,\cdots,h_s),$$

或

$$h_j=\sum_{i=1}^s\eta^{(i)}\omega_j^{(i)}\quad(1\leqslant j\leqslant s)$$

故由 (2.1) 式得

$$|\eta\omega_j-h_j|\leqslant\sum_{i=2}^s|\eta^{(i)}||\omega_j^{(i)}|\leqslant c(\mathscr{F}_s)|\eta|^{-\frac{1}{s-1}}.$$

命

$$1=\sum_{i=1}^{s}a_1\omega_i \quad 及 \quad n=\sum_{i=1}^{s}a_ih_i.$$

有

$$|\eta-n|=\left|\sum_{i=1}^{s}a_i\omega_i\eta-\sum_{i=1}^{s}a_ih_i\right|\leqslant\sum_{i=1}^{s}|a_i||\omega_i\eta-h_i|$$
$$\leqslant c(\mathscr{F}_s)|\eta|^{\frac{1}{s-1}},$$

即得

$$\left|\omega_j-\frac{h_j}{n}\right|\leqslant c(\mathscr{F}_s)|n|^{-1-\frac{1}{s-1}}(1\leqslant j\leqslant s). \tag{2.4}$$

当 $s>2$ 时, 经典方法只能证明存在无限多组整数 $(h_1,\cdots,h_s;n)$, 使 (2.4) 式成立. 但不能具体定出 $(h_1,\cdots,h_s;n)$ 的方法. 本书将指出, 寻求 $(h_1,\cdots,h_s;n)$ 的问题等价于寻求完实代数数域 $\mathscr{F}_s$ 中适合 (2.1) 式的单位 $\eta$ 的问题. 若知道完实代数数域 $\mathscr{F}_s$ 的一个独立单位组, 则可以由下面定理求出适合 (2.1) 式, 而且满足

$$|\eta_l|\to\infty(当 l\to\infty)$$

的单位贯 $(\eta_l)$ 来. 因而有无穷多组有理整数 $(h_1,\cdots,h_s;n)$ 适合 (2.4) 式.

**定理 2.1** 完实代数数域 $\mathscr{F}_s$, 中存在单位贯 $\eta_l(=\eta_l^{(1)})(l=1,2,\cdots)$, 其共轭数适合

$$|\eta_l^{(i)}|<e^{-(2l-1)c}\quad(2\leqslant i\leqslant s),$$

及

$$e^{-2c}|\eta_l^{(j)}|\leqslant|\eta_l^{(i)}|\leqslant e^{2c}|\eta_l^{(j)}|\quad(2\leqslant i,j\leqslant s),$$

此处 $c=c(\mathscr{F}_s)>0$.

**证** 命 $\varepsilon_1,\cdots,\varepsilon_{s-1}$ 为 $\mathscr{F}_s$ 的一组独立单位, 命

$$\xi^{(i)}=\varepsilon_1^{(i)l_1}\cdots\varepsilon_{s-1}^{(i)l_{s-1}}\quad(2\leqslant i\leqslant s)$$

及

$$c=\max_{2\leqslant i\leqslant s}\left(\sum_{j=1}^{s-1}|\log|\varepsilon_j^{(i)}||\right).$$

由于

$$\det(\log|\varepsilon_1^{(i)l_1}|)\neq 0\quad(2\leqslant i\leqslant s,1\leqslant j\leqslant s-1),$$

我们记线性方程组

$$\log|\xi^{(2)}|=\cdots=\log|\xi^{(s)}|=-2cl-1$$

的解为

$$l_1 = l_1^{(l)}, \cdots, l_{s-1} = l_{s-1}^{(l)}.$$

命

$$[l_i^{(l)}] = a_i^{(l)} \quad (1 \leqslant i \leqslant s-1),$$

此处 $[x]$ 表示 $x$ 的整数部分. 定义

$$\eta_l = \varepsilon_1^{a_1^{(l)}} \cdots \varepsilon_{s-1}^{a_{s-1}^{(l)}}.$$

则

$$\begin{aligned}
\log|\eta_l^{(i)}| &= \sum_{j=1}^{s-1} a_j^{(l)} \log|\varepsilon_j^{(i)}| \\
&\leqslant \sum_{j=1}^{s-1} l_j^{(l)} \log|\varepsilon_j^{(i)}| + \sum_{j=1}^{s-1} |\log|\varepsilon_j^{(i)}|| \\
&= \log|\xi^{(i)}| + \sum_{j=1}^{s-1} |\log|\varepsilon_j^{(i)}|| < -2cl + c \\
&= -(2l-1)c \quad (2 \leqslant i \leqslant s),
\end{aligned}$$

及

$$\begin{aligned}
|\log|\eta_l^{(i)}| - \log|\eta_l^{(j)}|| &\leqslant |\log|\xi^{(i)}| - \log|\xi^{(j)}|| \\
&+ \sum_{k=1}^{s-1}(|\log|\varepsilon_k^{(i)}|| + |\log|\varepsilon_k^{(j)}||) \leqslant 2c \quad (2 \leqslant i, j \leqslant s).
\end{aligned}$$

定理 2.1 得证.

由定理 2.1 可见, 欲得到满足 (2.4) 的整数组 $(h_1, \cdots, h_; n)$, 仅需 $c(\mathscr{F}_s)\log|n|$ 次初等运算.

## §3. 实分圆域与实 Dirichlet 域

(1) 命 $p \geqslant 5$ 为一素数, $r = \frac{1}{2}(p-1)$ 及 $1 = r - 1 = \frac{1}{2}(p-3)$, 完实分圆域 $\mathscr{R}_r = R\left(2\cos\frac{2\pi}{p}\right)$ 是一个 $r$ 次域, 它的一组整底是

$$\omega_1 = 2\cos\frac{2\pi}{p}, \omega_2 = 2\cos\frac{4\pi}{p}, \cdots, \omega_r = 2\cos\frac{2\pi r}{p}. \tag{3.1}$$

命 $\sigma$ 为 $(\omega_1,\cdots,\omega_r)$ 的轮换 (即 $\omega_1\to\omega_2,\cdots,\omega_r\to\omega_1$), 则 $\mathscr{R}_r$ 中的任何元素 $\eta(=\eta^{(1)})$, 经变换 $\sigma,\sigma^2,\cdots,\sigma^q$ 后, 得到 $\eta$ 的 $q$ 个共轭元素 $\eta^{(2)},\cdots,\eta^{(r)}$, 因此

$$S=\Omega'\Omega=pI-2M.$$

此处 $I$ 为单位方阵及 $M=(m_{ij}),m_{ij}=1$.

取 $\mathscr{R}_r$ 的一组独立单位 $\varepsilon_1,\cdots,\varepsilon_q$. 则由定理 2.1 构造单位

$$\eta=\varepsilon_1^{a_1}\cdots\varepsilon_q^{a_q}$$

满足

$$|\eta|>1,e^{-2c}|\eta^{(j)}|\leqslant|\eta^{(i)}|\leqslant e^{2c}|\eta^{(j)}|\quad(2\leqslant i,j\leqslant r),$$

此处 $c=c(\mathscr{R}_r)>0$. 若

$$\eta=\sum_{i=1}^{r}k_i\omega_i,$$

则由

$$(h_1,\cdots,h_r)=(k_1,\cdots,k_r)S=(k_1,\cdots,k_r)(pI-2M),$$

得

$$h_i=pk_i-2\sum_{j=1}^{r}k_j\quad(1\leqslant i\leqslant r),$$

由于

$$\sum_{i=1}^{r}\omega_i=-1,$$

所以

$$n=-\sum_{j=1}^{r}h_j=-\sum_{j=1}^{r}k_j.$$

故得联立丢番图逼近

$$\left|\frac{h_i}{n}-\omega_i\right|\leqslant c(\mathscr{R}_r)n^{-1-\frac{1}{q}}\quad(1\leqslant i\leqslant r). \tag{3.2}$$

悉知

$$p_l=\frac{\sin\dfrac{\pi}{p}g^{l+1}}{\sin\dfrac{\pi}{p}g^{l}}\quad(1\leqslant l\leqslant q)$$

为分圆域 $\mathscr{R}_r$ 的一组独立单位, 此处 $g$ 为 $\bmod p$ 的原根. 如果能用单位组

$$\omega_l=2\cos\frac{2\pi l}{p}\quad(1\leqslant l\leqslant q)$$

则更为方便, 但它们并不总是独立的, 但我们知道 $2\cos\frac{2\pi l}{p}(1\leqslant l\leqslant q)$ 为一组独立单位的充要条件是: (i) 2 为 mod $p$ 的原根或 (ii) 2 的次数为 $r$, 而且 $p\equiv 7($ mod 8)(见文献 [9]).

(2) 命 $p_1,\cdots,p_h$ 为 $h$ 个互不相同的素数, $m=2^h$ 及 $l=m-1$. 实 Dirichlet 域 $\mathscr{D}_m=R(\sqrt{p_1},\cdots,\sqrt{p_h})$ 为一个 $m$ 次代数数域. 命 $\varepsilon_1,\cdots,\varepsilon_l$ 为 $\mathscr{D}_m$ 的一组独立单位. 我们可以取诸 $\varepsilon_i$ 为诸 Pell 氏方程

$$x^2-p_i\cdots p_{i_k}y^2=\pm 4$$

的最小解 $\frac{x}{2}+\frac{\sqrt{p_{i_1}\cdots p_{i_k}}y}{2}$, 此处 $k\geqslant 1$ 及 $1\leqslant i_1<\cdots<i_k\leqslant h$ 为 $1,2,\cdots,h$ 的任意选择. 可以取

$$\varepsilon_i\left\{\begin{array}{l}\frac{x_i}{2}+\frac{\sqrt{d_i}y_i}{2}, x_i\equiv y_i\equiv 1(\mathrm{mod}2)\quad(1\leqslant i\leqslant\tau),\\ x_i+\sqrt{d_i}y_i\quad(\tau+1\leqslant i\leqslant l),\end{array}\right.$$

此处 $x_i$ 与 $y_i$ 为有理整数.

$\mathscr{D}_m$ 的整数为:

$$\omega_1=1,\omega_2=\varepsilon_1,\cdots,\omega_{\tau+1}=\varepsilon_\tau,\omega_{\tau+2}=\sqrt{d_{\tau+1}},\cdots,\omega_m=\sqrt{d_l}.$$

考虑 $l$ 个变换

$$(\sigma_{i_1},\cdots,i_k)\quad\sqrt{p_v}\to\left\{\begin{array}{ll}-\sqrt{p_v}, & 当v=i_j,\\ \sqrt{p_v}, & 当v\neq i_j,\end{array}\right.\ 1\leqslant j\leqslant k,$$

此处 $k\geqslant 1$ 及 $1\leqslant i_1<\cdots<i_k\leqslant h$ 为 $1,2,\cdots,h$ 的任意选择. 将这些变换作用于 $\mathscr{D}_m$ 的任何元素 $\eta(=\eta^{(1)})$, 则得其 $l$ 个共轭元素 $\eta^{(2)},\cdots,\eta^{(m)}$. 构造方阵

$$\Omega=(\omega_i^{(j)})\quad(1\leqslant i,j\leqslant m),$$

则

$$S=\Omega'\Omega=\begin{pmatrix}A&0\\0&B\end{pmatrix},$$

此处 $A=(a_{ij})(1\leqslant i,j\leqslant\tau+1)$ 及 $B=(b_{\mu\nu})(1\leqslant\mu,\nu\leqslant 1-\tau)$, 其中 $a_{11}=2^h,a_{ii}=2^{h-2}(x_{i-1}^2+d_{i-1}y_{i-1}^2)(2\leqslant i\leqslant\tau+1),a_{1j}=a_{j1}=2^{h-1}x_{j-1}(2\leqslant j\leqslant\tau+1),a_{ij}=2^{h-2}x_{i-1}x_{j-1}(2\leqslant i,j\leqslant\tau+1,i\neq j),b_{\mu\mu}=2^hd_{\tau+\mu}(1\leqslant\mu\leqslant l-\tau)$ 及 $b_{\mu\nu}=0(\mu\neq\nu)$.

按定理 2.1 构造 $\mathscr{D}_m$ 的单位

$$\eta=\varepsilon_1^{a_1}\cdots\varepsilon_l^{a_l},$$

使之满足

$$|\eta|>1,\quad e^{-2c}|\eta^{(j)}|\leqslant|\eta^{(i)}|\leqslant e^{2c}|\eta^{(j)}|\quad(2\leqslant i,j\leqslant m),$$

此处 $c=c(\mathscr{D}_m)>0$, 若

$$\eta=\sum_{i=q}^{m}k_i\omega_i,$$

则由 $(k_1,\cdots,k_m)S=(h_1,\cdots,h_m)$ 得

$$n=h_1=2^{h-1}\left(2k_1+\sum_{i=1}^{\tau}x_ik_{i+1}\right),$$

$$h_j=\begin{cases}2^{h-2}\left(2x_{j-1}k_1+d_{j-1}y_{j-1}^2k_j+\sum\limits_{i=2}^{\tau+1}x_{i-1}x_{j-1}k_i\right)\quad(2\leqslant j\leqslant\tau+1),\\ 2^hk_jd_{j-1}(\tau+2\leqslant j\leqslant m).\end{cases}$$

故得联立丢番图逼近

$$\left|\omega_i-\frac{h_i}{n}\right|\leqslant c(\mathscr{D}_m)n^{-1-\frac{1}{l}}\quad(1\leqslant i\leqslant m).$$

## §4. 一 致 分 布

本文用 $\boldsymbol{\gamma}=(\gamma_1,\cdots,\gamma_s)$ 表示实分量矢量, 而用 $\boldsymbol{m}=(m_1,\cdots,m_s)$ 表示整数分量矢量. 我们还引入记号 $||\boldsymbol{\gamma}||=\bar{\gamma}_1\cdots\bar{\gamma}_s,|\boldsymbol{\gamma}|=|\gamma_1,\cdots,\gamma_s|$ 及 $(\boldsymbol{\alpha},\boldsymbol{\beta})=\sum\limits_{i=1}^{s}\alpha_i\beta_l$($\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的矢量乘积). 现在我们证明 Erdös-Turan-Koksma 公式.

**定理 4.1** 命 $\frac{1}{6}>\eta>0$ 及 $h>\frac{1}{\eta}$ 为整数, 则对于任何 $\gamma\in G_s$ 皆有

$$\left|\frac{1}{n_l}N_{n_l}(\gamma)-|\gamma|\right|<2^{s+2}\varphi(n_l),$$

此处

$$\varphi(n_l)=\sum_{|m_i|\leqslant h}{}'\frac{1}{||\pi\boldsymbol{m}||}\left|\frac{1}{n}\sum_{j=1}^{n_l}e^{2\pi i(P_{n_l}(j).m)}\right|+\frac{\log^s 9h}{\eta h}+2^{s+1}\eta,$$

其中 $\Sigma'$ 表示在求和号中略去 $\boldsymbol{m}=\boldsymbol{0}=(0,\cdots,0)$ 一项.

**引理 4.1** 命 $r$ 为正整数, $\alpha,\beta$ 为实数及 $\Delta$ 满足

$$0<\Delta<\frac{1}{2}, \Delta\leqslant\beta-\alpha\leqslant 1-\Delta.$$

则存在以 1 为周期的函数 $\Psi(x)$, 满足以下条件:

(1) $\Psi(x)=1$, 当 $\alpha+\frac{1}{2}\Delta\leqslant x\leqslant\beta-\frac{1}{2}\Delta$.

(2) $0\leqslant\Psi(x)\leqslant 1$, 当 $\alpha-\frac{1}{2}\Delta\leqslant x\leqslant\alpha+\frac{1}{2}\Delta$
及

$$\beta-\frac{1}{2}\Delta\leqslant x\leqslant\beta+\frac{1}{2}\Delta.$$

(3) $\Psi(x)=0$, 当 $\beta+\frac{1}{2}\Delta\leqslant x\leqslant 1+\alpha-\frac{1}{2}\Delta$.

(4) $\Psi(x)$ 有 Fourier 展开

$$\Psi(x)=\beta-\alpha+\sum{}' C(m)e^{2\pi imx},$$

此处

$$|C(m)|\leqslant\min\left(\beta-\alpha,\frac{1}{\pi|m|},\left(\frac{1}{\pi|m|}\right)^{r+1}\left(\frac{r}{\Delta}\right)^{r}\right).$$

证明见文献 [15]

**引理 4.2** 命 $0<\delta\leqslant\varphi(n_l)$. 若当 $\delta\leqslant\gamma_i\leqslant(1-\delta(1\leqslant i\leqslant s)$ 时, 一致分布点集贯 $(P_{n_l}(j))(n_1<n_2<\cdots)$ 满足

$$\left|\frac{1}{n_l}N_{n_l}\boldsymbol{\gamma}-|\boldsymbol{\gamma}|\right|<\varphi(n_l),$$

则对于任何 $\gamma\in G_s$ 皆有

$$\left|\frac{1}{n_l}N_{n_l}(\boldsymbol{\gamma})-|\boldsymbol{\gamma}|\right|<2^{s+2}\varphi(n_l).$$

**证** 为简单起见, 今后我们皆略去 $n_l$ 与 $h^{(l)}$ 的指标 $l$.

(1) 命 $\delta\leqslant\alpha_i<\beta_i\leqslant 1-\delta(1\leqslant i\leqslant s)$ 及 $N_n(\alpha,\boldsymbol{\beta})$ 表示 $P_n(j)(1\leqslant j\leqslant n)$ 满足

$$\alpha_i\leqslant x_i^{(n)}(j)<\beta_i\quad(1\leqslant i\leqslant s)$$

的个数. 则易知

$$\left|\frac{1}{n}N_n(\boldsymbol{\alpha},\boldsymbol{\beta})-|\boldsymbol{\beta}-\boldsymbol{\alpha}|\right|<2^s\varphi(n).$$

(2) 命 $\mathscr{D}$ 表示区域

$$\delta\leqslant x_i<1-\beta\quad(1\leqslant i\leqslant s)$$

及 $\overline{\mathscr{D}}$ 表示 $\mathscr{D}$ 关于 $G_s$ 的余集, 则 $\overline{\mathscr{D}}$ 的测度 $\leqslant 2^s\delta \leqslant 2^s\varphi(n)$. 由 (1) 可知落人 $\mathscr{D}$ 的点 $P_n(j)$ 的个数为

$$(1-2\delta)^s n + \vartheta 2^s n\varphi(n).$$

$\vartheta$ 不一定表示相同的数, 但 $|\vartheta| \leqslant 1$, 所以属于 $\overline{\mathscr{D}}$ 的点 $P_n(j)$ 的个数不超过

$$(1-(1-2\delta)^s)n + 2^s\varphi(n)n \leqslant 2^{s+1}\varphi(n)n.$$

(3) 综合 (1)、(2) 可知, 对于任何 $\gamma \in G_s$ 皆有

$$\left|\frac{1}{n_l}N_n(\gamma) - |\gamma|\right| < (2^s + 2^s + 2^{s+1})\varphi(n) = 2^{s+2}\varphi(n).$$

引理 4.2 证完.

**定理 4.1 的证明** 由引理 4.2 可知, 只要在条件 $3\eta \leqslant \gamma_i \leqslant 1-3\eta (1 \leqslant i \leqslant s)$ 之下证明

$$\left|\frac{1}{n}N_n(\gamma) - |\gamma|\right| < \varphi(n)$$

即可.

引入函数

$$G_x(y) = \begin{cases} 1, 当 0 \leqslant y < x, \\ 0, 当 x \leqslant y < 1. \end{cases}$$

则得

$$\frac{1}{n}N_n(\boldsymbol{x}) = \frac{1}{n}\sum_{j=1}^{n} G_{x_1}(x_1^{(n)}(j)) \cdots G_{x_s}(x_s^{(n)}(j)). \tag{4.1}$$

当 $3\eta \leqslant x \leqslant 1-3\eta$ 时, 按引理 4.1 构造两个辅助函数 $G_x^{(1)}(y)$ 与 $G_x^{(2)}(y)$.$G_x^{(1)}(y)$ 满足

(i) $G_x^{(1)}(y) = 1$, 当 $2\eta \leqslant y \leqslant x-\eta$;

(ii) $0 \leqslant G_x^{(1)}(y) \leqslant 1$, 当 $\eta \leqslant y \leqslant 2\eta$ 及 $x-\eta \leqslant y \leqslant x$;

(iii) $G_x^{(1)}(y) = 0$, 当 $x \leqslant y \leqslant 1+\eta$;

(iv) $G_x^{(1)}(y)$ 有 Fourier 展开

$$G_x^{(1)}(y) = x - 2\eta + \sum{}' C_2(m)e^{2\pi i m y},$$

其中

$$|C_1(m)| \leqslant \min\left(x-2\eta, \frac{1}{\pi|m|}, \frac{1}{\eta\pi^2 m^2}\right).$$

$G_x^{(2)}(y)$ 满足

(i)′ $G_x^{(2)}(y) = 1$, 当 $-\eta \leqslant y \leqslant x$;

(ii)′ $0 \leqslant G_x^{(2)}(y) \leqslant 1$, 当 $-2\eta \leqslant y \leqslant -\eta$ 及 $x \leqslant y \leqslant x+\eta$;
(iii)′ $G_x^{(2)}(y) = 0$, 当 $x+\eta \leqslant y \leqslant 1-2\eta$;
(iv)′ $G_x^{(2)}(y)$ 有 Fourier 展开

$$G_x^{(2)}(y) = x + 2\eta + {\sum}' C_2(m) e^{2\pi i m y},$$

其中

$$|C_2(m)| \leqslant \min\left(x+2\eta, \frac{1}{\pi|m|}, \frac{1}{\eta\pi^2 m^2}\right).$$

由 (iv) 可知

$$\begin{aligned} G_s^{(1)}(y) =& x - 2\eta + {\sum_{|m|\leqslant h}}' C_1(m) e^{2\pi i m y} + \vartheta \sum_{|m|>h} \frac{1}{\eta\pi^2 m^2} \\ =& x - 2\eta + {\sum_{|m|\leqslant h}}' C_1(m) e^{2\pi i m y} + \frac{2\vartheta}{\pi^2 \eta h}. \end{aligned}$$

其中 $\sum\limits_{m>h} \dfrac{1}{m^2} \leqslant \displaystyle\int_h^\infty \frac{dt}{t^2} = h^{-1}$. 在上式中置 $x = x_i$ 及 $y = y_i$, 并置 $C_0^{(1)} = x_i - 2\eta (1 \leqslant i \leqslant s)$. 则得 $s$ 个方程, 分别将左端相乘与右端相乘, 则得

$$\begin{aligned} G_{x_1}^{(1)}(y_1) \cdots G_{x_s}^{(1)}(y_s) =& \sum_{|m_i|\leqslant h} C_1(m_1) \cdots C_1(m_s) e^{2\pi i (m_1 y_1 + \cdots + m_s y_s)} \\ & + \frac{2\vartheta}{\pi^2 \eta h} \left(1 + \sum_{|m|\leqslant h} |C_1(m)|\right)^s = (x_1 - 2\eta) \cdots (x_s - 2\eta) \\ & + {\sum_{|m_i|\leqslant h}}' C_1(m_1) \cdots C_1(m_s) e^{2\pi i (m_1 y_1 + \cdots + m_s y_s)} + \frac{\vartheta \log^s 9h}{\eta h}, \end{aligned}$$

其中

$$1 + \sum_{|m|\leqslant h} |C_1(m)| \leqslant 2 + \frac{2}{\pi} + \frac{2}{\pi} \int_1^h \frac{dt}{t} < 2 + \log h < \log 9h.$$

所以

$$\begin{aligned} G_{x_1}^{(1)}(y_1) \cdots G_{x_s}^{(1)}(y_s) =& x_1 \cdots x_s + {\sum_{|m_i|\leqslant h}}' C_1(m_1) \cdots C_1(m_s) \\ & e^{2\pi i (m_1 y_1 + \cdots + m_s y_s)} + \vartheta \left(\frac{\log^s 9h}{\eta h} + 2^{s+1}\eta\right). \end{aligned} \tag{4.2}$$

类似地, 有

$$\begin{aligned}&G_{x_1}^{(2)}(y_1)\cdots G_{x_s}^{(2)}(y_s)\\&=x_1\cdots x_s+\sum_{|m_i|\leqslant h}{}' C_2(m_1)\cdots C_2(m_s)e^{2\pi i(m_1y_1+\cdots+m_sy_s)}\\&\quad+\vartheta\left(\frac{\log^s 9h}{\eta h}+2^{s+1}\eta\right).\end{aligned}\tag{4.3}$$

由 $G_x^{(1)}(y)$ 与 $G_x^{(2)}(y)$ 的定义可知,

$$\begin{aligned}G_{x_1}^{(1)}(y_1)\cdots G_{x_s}^{(1)}(y_s)\leqslant&G_{x_1}(y_1)\cdots G_{x_s}(y_s)\\\leqslant&G_{x_1}^{(2)}(y_1)\cdots G_{x_s}^{(2)}(y_s).\end{aligned}\tag{4.4}$$

故由 (4.1)、(4.2)、(4.3)、(4.4) 式即得定理 2.2.

## §5. 定理 1 的证明

**定理 5.1** 若对于任何矢量 $\boldsymbol{m}\neq\boldsymbol{0}$ 皆有不等式

$$\langle(\boldsymbol{m},\gamma)\rangle>b\|\boldsymbol{m}\|^{-a},\tag{5.1}$$

此处 $a,b$ 为满足 $s+1\geqslant a\geqslant 1$ 及 $1\geqslant b>0$ 的常数, 则贯

$$P(j)=(\{\gamma_1 j\},\cdots,\{\gamma_s j\})(j=1,2,\cdots)$$

有偏差

$$\varphi(n)=c(a,b,s)n^{-1+2s(a-1)}(\log 3n)^{1+s\delta_{1,a}},$$

此处 $\delta_{\alpha,\beta}$ 表示 Kronecker 符号

**引理 5.1** 命 $\delta$ 为任意实数. 则

$$\left|\sum_{j=1}^{n}e^{2\pi i\delta j}\right|\leqslant\min\left(n,\frac{1}{2\langle\delta\rangle}\right).$$

**证** 若 $\delta$ 为非整数, 则

$$\begin{aligned}\left|\sum_{j=1}^{n}e^{2\pi i\delta j}\right|=&\left|\frac{e^{2\pi i\delta(n+1)}-e^{2\pi i\delta}}{e^{2\pi i\delta}-1}\right|\\\leqslant&\frac{1}{|\sin\pi\delta|}\leqslant\frac{1}{2\langle\delta\rangle}.\end{aligned}$$

故得引理 5.1.

**引理 5.2**　命 $g(\boldsymbol{m})$ 表示 $\boldsymbol{m}$ 的非负函数, 则

$$\sum_{|m_i|\leqslant h}{}'\frac{g(\boldsymbol{m})}{||\boldsymbol{m}||}\leqslant\sum_{l=0}^{s}\sum_{i}\frac{1}{h^l}\sum_{m_{i_{l+1}}=1}^{h}\cdots\sum_{m_{i_s}=1}^{h}\frac{1}{m_{i_{l+1}}^2\cdots m_{i_s}^2}\times\sum_{|k_{i_1}|\leqslant h+1}\cdots\sum_{|k_{i_l}|\leqslant h+1}\sum_{|k_{i_{l+1}}|\leqslant m_{i_{l+1}}}\cdots\sum_{|k_{i_s}|\leqslant m_{i_s}}{}'g(\boldsymbol{k}),$$

此处 $\sum\limits_{i}$ 表示通过 $(1,2,\cdots,s)$ 所有的置换 $\boldsymbol{i}=(i_1,\cdots,i_s)$ 求和.

**证**　由于

$$\begin{aligned}\sum_{|m|\leqslant h}\frac{g(m)}{m}&=g(0)+\sum_{m=1}^{h}\frac{1}{m}(g(m)+g(-m))\\&=g(0)+\sum_{m=1}^{h}\left(\frac{1}{m}-\frac{1}{m+1}\right)\sum_{1\leqslant k\leqslant m}(g(k)+g(-k))\\&\quad+\frac{1}{h+1}\sum_{|k|\leqslant h+1}{}'g(k)\\&\leqslant\sum_{m=1}^{h}\frac{1}{m^2}\sum_{|k|\leqslant m}g(k)+\frac{1}{h}\sum_{|k|\leqslant h+1}g(k),\end{aligned}$$

所以

$$\begin{aligned}&\sum_{|m_i|\leqslant h}{}'\frac{g(\boldsymbol{m})}{||\boldsymbol{m}||}\\&\leqslant\sum_{|m_i|\leqslant h}{}'\frac{1}{\bar{m}_1\cdots\bar{m}_{s-1}}\left(\sum_{m_s=1}^{h}\frac{1}{m_s^2}\sum_{|k_s|\leqslant m_s}g(m_1,\cdots,m_{s-1},k_s)\right.\\&\quad\left.+\frac{1}{h}\sum_{|k_s|\leqslant h+1}g(m_1,\cdots,m_{s-1},k_s)\right)\\&\quad+\sum_{|m_i|\leqslant h}\frac{1}{\bar{m}_1\cdots\bar{m}_{s-1}}\left(\sum_{m_s=1}^{h}\frac{1}{m_s^2}\sum_{|k_s|\leqslant m_s}{}'g(m_1,\cdots,m_{s-1},k_s)\right.\\&\quad\left.+\frac{1}{h}\sum_{|k_s|\leqslant h+1}{}'g(m_1,\cdots,m_{s-1},k_s)\right)\leqslant\cdots\\&\leqslant\sum_{l=0}^{s}\sum_{i}\frac{1}{h^l}\sum_{m_{i_{l+1}}=1}^{h}\cdots\sum_{m_{i_s}=1}^{h}\frac{1}{m_{i_{l+1}}^2\cdots m_{i_s}^2}\\&\quad\times\sum_{|k_{i_1}|\leqslant h+1}\cdots\sum_{|k_{i_l}|\leqslant h+1}\sum_{|k_{i_{l+1}}|\leqslant m_{i_{l+1}}}\cdots\sum_{|k_{i_s}|\leqslant m_{i_s}}{}'g(\boldsymbol{k}).\end{aligned}$$

引理 5.2 证完.

**引理 5.3** 命 $\boldsymbol{m} \neq 0$ 及 $Q = [2^{sa}||\boldsymbol{m}||^a b^{-1}] + 1$. 若对于任何 $\boldsymbol{m} \neq \boldsymbol{0}$ 皆有 (5.1) 式, 则在任何区间 $(P, P + Q^{-1}]$ 中最多只包含一个点 $(\boldsymbol{k}, \boldsymbol{\gamma}) = \sum_{i=1}^{s} k_i \gamma_i$, 此处 $\boldsymbol{k}$ 为有整数分量的矢量, 其分量 $k_i$ 满足 $|k_i| \leqslant |m_i| (1 \leqslant i \leqslant s)$.

**证** 若在区间 $(P, P+Q^{-1}]$ 中包含两个点 $(\boldsymbol{k}', \boldsymbol{\gamma})$ 与 $(\boldsymbol{k}'', \boldsymbol{\gamma})$, 此处 $\boldsymbol{k}' \neq \boldsymbol{k}''$, $|k_i'| \leqslant |m_i|$ 及 $|k_i''| \leqslant |m_i| (1 \leqslant i \leqslant s)$, 则

$$\langle(\boldsymbol{k}' - \boldsymbol{k}'' \boldsymbol{\gamma})\rangle \leqslant Q^{-1}.$$

另一方面, 由 (5.1) 式可得

$$\begin{aligned}\langle(\boldsymbol{k}' - \boldsymbol{k}'', \boldsymbol{\gamma})\rangle >& b||\boldsymbol{k}' - \boldsymbol{k}''||^{-a} \\ \geqslant& 2^{-sa} b ||\boldsymbol{m}||^{-a} > Q^{-1}.\end{aligned}$$

故得出矛盾.

引理 5.3 得证.

**引理 5.4** 命 $\boldsymbol{m} \neq \boldsymbol{0}$ 及 $Q = [2^{sa}||\boldsymbol{m}||^a b^{-1}] + 1$. 若对于任何 $\boldsymbol{m} \neq \boldsymbol{0}$ 皆有 (5.1) 式, 则

$$\sum_{|k_i| \leqslant m_{|i|}}{}' \frac{1}{\langle(\boldsymbol{k}, \boldsymbol{\gamma})\rangle} \leqslant 4Q \log 3Q.$$

**证** 将区间 (0,1] 分成 $Q$ 个子区间

$$I_j = \left(\frac{j}{Q}, \frac{j+1}{Q}\right] \quad (j = 0, 1, \cdots, Q-1).$$

由 (5.1) 式可知在 $I_0$ 中不能包含点 $(\boldsymbol{k}, \boldsymbol{\gamma})$, 此处 $\boldsymbol{k} \neq \boldsymbol{0}$ 及 $|k_i| \leqslant |m_i| (1 \leqslant i \leqslant s)$. 由引理 5.3 可知对任何 $I_j$ 中最多只包含一个点 $(\boldsymbol{k}, \boldsymbol{\gamma})$, 此处 $j \geqslant 1$. 因此

$$\sum_{|k_i| \leqslant m_{|i|}}{}' \frac{1}{\langle(\boldsymbol{k}, \boldsymbol{\gamma})\rangle} \leqslant 4 \sum_{i=1}^{Q-1} \frac{Q}{j} \leqslant 4Q \log 3Q.$$

引理 5.4 证完.

**引理 5.5** 命 $h \geqslant 2$ 为整数. 若对于任何 $\boldsymbol{m} \neq \boldsymbol{0}$ 皆有 (5.1) 式, 则

$$\sum_{|m_i| \leqslant h}{}' \frac{1}{||\boldsymbol{m}|| \langle(\boldsymbol{m}, \boldsymbol{\gamma})\rangle} \leqslant c(a, b, s) h^{s(a-1)} (\log h)^{1+s\delta_{1,a}}$$

**证** 由引理 5.2 与引理 5.4 得

$$\begin{aligned}
&\sum_{|m_i|\leqslant h}{}' \frac{1}{||\boldsymbol{m}||\langle(\boldsymbol{m},\boldsymbol{\gamma})\rangle} \leqslant \sum_{l=0}^{s}\sum_{i}\frac{1}{h^l}\sum_{m_{i_{l+1}}=1}^{h}\cdots\sum_{m_{i_s}=1}^{h}\frac{1}{m_{i_{l+1}}^2\cdots m_{i_s}^2}\\
&\times\sum_{|k_{i_1}|\leqslant h+1}\cdots\sum_{|k_{i_l}|\leqslant h+1}\sum_{|k_{i_{l+1}}|\leqslant m_{i_{l+1}}}\cdots\sum_{|k_{i_s}|\leqslant m_{i_s}}{}'\frac{1}{\langle(\boldsymbol{k},\boldsymbol{\gamma})\rangle}\\
&\leqslant\sum_{l=0}^{s}\sum_{i}\frac{1}{h^l}\sum_{m_{i_{l+1}}=1}^{h}\cdots\sum_{m_{i_s}=1}\frac{1}{m_{i_{l+1}}^2\cdots m_{is}^2}c(a,b,s)h^{la}(m_{i_{l+1}}\cdots m_{is})^a\log h\\
&\leqslant c(a,b,s)h^{s(a-1)}(\log h)^{1+\delta_{1.a}}
\end{aligned}$$

引理 5.5 证完.

**定理 5.1 的证明** 由引理 5.1 与引理 5.5 可知

$$\begin{aligned}
&\sum_{|m_i|\leqslant h}{}'\frac{1}{||\boldsymbol{m}||}\left|\frac{1}{n}\sum_{j=1}^{n}e^{2\pi i(m,\gamma)j}\right|\leqslant\frac{1}{n}\sum_{|m_i|\leqslant h}{}'\frac{1}{2||\boldsymbol{m}||\langle(\boldsymbol{m},\boldsymbol{\gamma})\rangle}\\
&\leqslant c(a,b,s)n^{-1}h^{s(a-1)}(\log h)^{1+s\delta_{1.a}}
\end{aligned}$$

取 $\eta=\dfrac{1}{7n}$ 及 $h=8n^2$, 故由定理 4.1 即得定理 5.1.

由于 $1,\omega_1,\cdots,\omega_q$ 在有理数域 $R$ 上线性独立, 所以定理 1 显然是引理 1.1 与定理 5.1 以及定理 5.2 的推论.

## §6. 定理 2 的证明

我们用 $P_M^s$ 表示边平行于座标轴, 体积不超过 $M$ 的 $s$ 维平行多面体. 我们还用 $\boldsymbol{u}=(u_0,\cdots,u_s)$ 与 $\boldsymbol{h}=(1,h_1,\cdots,h_s)$ 表示 $(s+1)$ 维整数分量的矢量.

**定理 6.1** 命 $n>1$ 为整数及 $M\geqslant 1$. 又命 $\boldsymbol{a}=(a_1,\cdots,a_s)$ 为整数分量的矢量. 若同余式

$$(\boldsymbol{a},\boldsymbol{m})=\sum_{i=1}^{s}a_im_i\equiv 0(\mathrm{mod}n) \tag{6.1}$$

在区域

$$||\boldsymbol{m}||\leqslant M,\boldsymbol{m}\neq 0 \tag{6.2}$$

中无解, 则点集

$$\left(\left\{\frac{a_1j}{n}\right\},\cdots,\left\{\frac{a_sj}{n}\right\}\right)\quad(1\leqslant j\leqslant n)$$

有偏差

$$\varphi(n)=c(s,\varepsilon)M^{-1+\varepsilon}.$$

**定理 6.2**　命

$$\left|\frac{h_i}{n}-\gamma_i\right| \leqslant dn^{-1-\frac{1}{s}} \quad (1 \leqslant i \leqslant s) \tag{6.3}$$

为诸 $\gamma_i$ 的联立有理逼近, 此处 $d>0$ 为一常数. 若对于任何 $\boldsymbol{m} \neq \mathbf{0}$ 皆有 (5.1) 式, 则存在常数 $c(a,b,d,s)(<1)$ 使同余式

$$(\boldsymbol{h},\boldsymbol{u})=u_0+\sum_{i=1}^{s} h_l u_l \equiv 0(\bmod n) \tag{6.4}$$

在区域

$$||\boldsymbol{u}|| \leqslant c(a,b,d,s)n^{(1+\frac{1}{s})/(a+1)}, \boldsymbol{u} \neq \mathbf{0} \tag{6.5}$$

中无解

**引理 6.1**　命 $l \geqslant 1$ 为整数, 则 $s$ 维区域

$$||\boldsymbol{m}||<lM$$

可以被不超过 $c(\varepsilon)^s l^{1+\varepsilon} M^{\varepsilon}$ 个 $P_M^s$ 型的平行多面体所覆盖.

**证**　取

$$c(\varepsilon)=2^{2+\varepsilon}\sum_{j=0}^{\infty}(j^{-(1+\varepsilon)}+2^{-\varepsilon j}).$$

(1) 对于 $s=1$, 因区域 (6.6) 就是区间 $(-lM, lM)$, 所以它可以被不超过

$$\frac{2lM}{M}=2l$$

个型为 $[c,c+M]$ 的区间所覆盖, 此处 $c$ 为实数, 所以引理 6.1 对于 $s=1$ 成立.

(2) 我们假定 $k$ 为正整数, 而且引理 6.1 对于 $s=1,\cdots,k$ 成立. 现证明引理对于 $s=k+1$ 亦成立.

用超平面

$$m_{k+1}=0,\pm 2^i l \quad (i=0,1,\cdots,[\log_2 M])$$

将区域

$$\bar{m}_1\cdots\bar{m}_{k+1}<lM \tag{6.7}$$

分为 $2[\log_2 M]+3$ 片

(i) $m_{k+1}=j, j\leqslant l$

(ii) $2^i l<|m_{k+1}|\leqslant 2^{i+1}l(i=0,1,\cdots[\log_2 M])$.

(3) 假定 $m_{k+1}=j$. 则

$$\bar{m}_1 \cdots, \bar{m}_k < \frac{lM}{\bar{j}} < \left(\left[\frac{l}{\bar{j}}\right] + 1\right) M.$$

故由归纳法假定可知, 上述 $k$ 维区域可以被不超过

$$Q = c(\varepsilon)^k \left(\left[\frac{l}{\bar{j}}\right] + 1\right)^{1+\varepsilon} M^{\varepsilon}.$$

个 $P_M^k$ 型的平行多面体所覆盖, 以这些 $P_M^k$ 为底, 以 1 为高作 $P_M^{k+1}$, 则子区域 $m_{k+1} = j$ 可以被不超过 $Q$ 个 $P_M^{k+1}$ 型的平行多面体所覆盖. 因此由 (i) 定义的一片区域可以被不超过

$$2\sum_{j=0}^{l} c(\varepsilon)^k \left(\left[\frac{l}{j} + 1\right]\right)^{1+\varepsilon} M^{\varepsilon}. \tag{6.8}$$

个型 $P_m^{k+1}$ 的平行多面体所覆盖.

(4) 考虑 (6.7) 式的子区域

$$2^i l < m_{k+1} \leqslant 2^{i+1} l, \tag{6.9}$$

则当 $m_{k+1} = 2^i l + 1$ 时, 有

$$\bar{m}_1 \cdots \bar{m}_k < \frac{M}{2^i}.$$

故由归纳法假定可知上面区域可以被不超过

$$c(\varepsilon)^k \left(\frac{M}{2^i}\right)^{\varepsilon} \tag{6.10}$$

个 $P_{M/2^i}^k$ 型的平行多面体所覆盖. 以 $P_{M/2^i}^k$ 为底, 以 $2^i$ 为高, 作多面体 $P_M^{k+1}$. 因 $2^i l + 2^i l + 1 > 2^{i+1} l$, 所以区域 (6.9) 可以被不超过 $c(\varepsilon)^k l(M/2^i)^{\varepsilon}$ 个 $P_M^{k+I}$ 型的平行多面体所覆盖. 因此由 (ii) 定义的区域可以被不超过

$$2c(\varepsilon)^k \sum_{i=0}^{[\log_2 M]} l \left(\frac{M}{2^i}\right)^{\varepsilon} \tag{6.11}$$

个 $P_M^{k+1}$ 型的平行多面体所覆盖.

(5) 由 (6.8) 与 (6.11) 式可知, 区域 (6.7) 可以被不超过

$$c(\varepsilon)^k l^{1+\varepsilon} M^{\varepsilon} \sum_{j=0}^{\infty} \left(\frac{2^{2+\varepsilon}}{j^{1+\varepsilon}} + \frac{2}{2^{\varepsilon j}}\right) \leqslant c(\varepsilon)^{k+1} l^{+\varepsilon} M^{\varepsilon}$$

个 $P_M^{k+1}$ 型的平行多面体所覆盖, 故由归纳法即得引理 6.1.

引理 6.1 中的 $c(\varepsilon)^s l^{l+\varepsilon} M^\varepsilon$ 可以换为 $c(s) l \log^{s-1} 3lM$(见文献 [6]).

**引理 6.2** 命 $T_M^l$ 表示同余式 (6.1) 在区域 (6.6) 中的解数. 若 (6.1) 式在区域 (6.2) 中无解, 则

$$T_M^l \leqslant c(\varepsilon)^s l^{1+\varepsilon} M^\varepsilon.$$

**证** 由引理 6.1 可知, 只要证明同余式 (6.1) 在任何 $P_M^s$ 型的平行多面体中最多只有一个解即可. 假定同余式 (6.1) 在某 $P_M^s$ 中有两个解 $\boldsymbol{m}'$ 与 $\boldsymbol{m}''$, 此处 $\boldsymbol{m}' \neq \boldsymbol{m}''$. 命 $\boldsymbol{m} = \boldsymbol{m}' - \boldsymbol{m}''$. 则 $||\boldsymbol{m}|| \leqslant M$ 及

$$(\boldsymbol{a}, \boldsymbol{m}) = (\boldsymbol{a}, \boldsymbol{m}') - (\boldsymbol{a}, \boldsymbol{m}'') \equiv 0 (\bmod n).$$

故得矛盾.

引理 6.2 得证

**定理 6.1 的证明** 取 $3\delta s = \varepsilon$. 则由引理 6.2 可知

$$\begin{aligned}
&\sum_{|m_i| \leqslant h}{}' \frac{1}{||\boldsymbol{m}||} \left| \frac{1}{n} \sum_{j=1}^{n} e^{2\pi i (\boldsymbol{a}, \boldsymbol{m}) j/n} \right| \\
= &\sum_{\substack{|m_i| \leqslant h \\ (\boldsymbol{a}, \boldsymbol{m}) \equiv 0 (\bmod n)}}{}' \frac{1}{||\boldsymbol{m}||} \leqslant \sum_{l=1}^{h^s} \frac{(T^{l+1} - T_M^l)}{lM} \\
= &\frac{1}{M} \sum_{t=1}^{h^s} T_M^{l+1} \left( \frac{1}{l} - \frac{1}{l+1} \right) + \frac{T_M^{h^s+1}}{(h^s+1)M} \\
\leqslant &c(\delta)^s M^{-1+\delta} h^{s\delta} \quad (T_M^1 = 0).
\end{aligned}$$

取 $\frac{1}{M}$ 及 $h = 7([M]+1)^2$. 则由定理 4.1 即得定理 6.1.

**定理 6.2 的证明** 命 $\boldsymbol{u} \neq \boldsymbol{0}$ 为同余式 (6.4) 的解. 若 $u_i = 0 (1 \leqslant i \leqslant s)$, 则 $u_0 \neq 0$. 由 (6.4) 式可知 $u_0 \equiv 0 (\bmod n)$. 所以 $||\boldsymbol{u}|| \geqslant n$. 从而 $\boldsymbol{u}$ 不属于区域 (6.5). 因此我们可以假定 $(u_1, \cdots, u_s) \neq (0, \cdots, 0)$. 若

$$\bar{u}_1 \cdots \bar{u}_s \geqslant \left( \frac{b}{2ds} \right)^{\frac{1}{a+1}} n^{(1+\frac{1}{s})/(a+1)},$$

则

$$||\boldsymbol{u}|| \geqslant \left( \frac{b}{2ds} \right)^{\frac{1}{a+1}} n^{(1+\frac{1}{s})/(a+1)},$$

故得定理. 现再假定

$$\bar{u}_1 \cdots \bar{u}_s < \left( \frac{b}{2ds} \right)^{\frac{1}{a+1}} n^{(1+\frac{1}{s})/(a+1)}.$$

因

$$\langle \alpha - \beta \rangle \geqslant \langle \alpha \rangle - \langle \beta \rangle,$$

由 (5.1) 与 (6.3) 式得

$$\begin{aligned}\frac{|u_0|}{n} &\geqslant \left\langle \frac{u_0}{n} \right\rangle = \left\langle \frac{1}{n} \sum_{i=1}^{n} h_i u_i \right\rangle \\ &\geqslant \left\langle \sum_{i=1}^{s} \gamma_i u_i \right\rangle - \left\langle \sum_{i=1}^{s} \left( \frac{h_i}{n} - \gamma_i \right) u_i \right\rangle \\ &\geqslant \frac{b}{(\bar{u}_1 \cdots \bar{u}_s)^a} - \frac{ds}{n^{1+\frac{1}{s}}} (\bar{u}_1 + \cdots \bar{u}_s).\end{aligned}$$

所以

$$|u_0| \bar{u}_1 \cdots \bar{u}_s \geqslant \frac{1}{2} (2ds)^{\frac{a-1}{a+1}} b^{\frac{2}{a+1}} n^{(1+\frac{1}{s})/(a+1)}.$$

定理 6.2 证完.

显然, 定理 2 是引理 1.1、定理 6.1 和定理 6.2 的推论.

## §7. 定理 3 的证明

我们用记号

$$I(f) = \int_0^1 \cdots \int_0^1 f(x_1, \cdots, x_s) dx_1 \cdots dx_s$$

及

$$P_n(f) = I(f) - \frac{1}{(2n+1)^6} \sum_{j=-nl}^{nl} \mu_{n,l,j} f(j\gamma).$$

**定理 7.1** 若对于任何 $\boldsymbol{m} \neq \mathbf{0}$ 皆有 (5.1) 式, 则

$$\sup_{f \in E_s^a(C)} |P_n(f)| \leqslant C \cdot c(a, b, a, s) n^{-a+sa(a-1)/(a-1)} (\log 3n)^{a+sa\delta_{1,a}}.$$

**证** 因 $\left| \dfrac{\sin h\pi\delta}{h \sin \pi\delta} \right| \leqslant 1,$

此处 $h$ 为正整数, 而 $\delta$ 为非整数的实数.

$$\frac{1}{(2n+1)^l} \sum_{l=-nl}^{nl} \mu_{n,l,j} f(l\gamma)$$

$$
\begin{aligned}
&=\frac{1}{(2n+1)^l}\sum C(\boldsymbol{m})\sum_{l=-nl}^{nl}\mu_{n,l,j}e^{2\pi i(m,\gamma)i}\\
&=C(\mathbf{0})+\frac{1}{(2n+1)^l}\sum{}'C(\boldsymbol{m})\left(\sum_{j=-n}^{n}e^{2\pi i(m,\gamma)j}\right)^l\\
&=C(\mathbf{0})+\sum{}'C(\boldsymbol{m})\left(\frac{\sin(2n+1)\pi(\boldsymbol{m},\boldsymbol{\gamma})}{(2n+1)\sin\pi(\boldsymbol{m},\boldsymbol{\gamma})}\right)^l,
\end{aligned}
$$

及

$$C(\mathbf{0})=I(f),$$

故得

$$
\begin{aligned}
\sup_{f\in E_s^\alpha(C)}|P_n(f)|\leqslant C\sum{}'\frac{1}{\|\boldsymbol{m}\|^\alpha}\left|\frac{\sin(2n+1)\pi(\boldsymbol{m},\boldsymbol{\gamma})}{(2n+1)\sin\pi(\boldsymbol{m},\boldsymbol{\gamma})}\right|^\alpha\\
=C\left(\sum\nolimits_1+\sum\nolimits_s\right),
\end{aligned}
\tag{7.1}
$$

此处 $\Sigma_1$ 表示满足 $|m_i|\leqslant n^{\frac{\alpha}{\alpha-1}}(1\leqslant i\leqslant s)$ 的 $\boldsymbol{m}$ 求和, 而 $\Sigma_2$ 则表示其余部分.

当 $\alpha>1, a_l>0$ 及 $\sum\limits_i a_l<\infty$ 时有

$$\sum_i a_i^a=\sum_i\left[\frac{a_i}{\sum\limits_j a_j}\right]^a\left(\sum_k a_k\right)^a\leqslant\sum_i\frac{a_i}{\sum\limits_j a_j}\left(\sum_k a_k\right)^a=\left(\sum_k a_k\right)^a.$$

由引理 5.1 与引理 5.5 可知

$$
\begin{aligned}
\sum\nolimits_1&\leqslant\frac{1}{2^\alpha(2n+1)^\alpha}\sum_{|m_i|\leqslant n^{\frac{\alpha}{\alpha-1}}}{}'\frac{1}{\|\boldsymbol{m}^\alpha\|\langle(\boldsymbol{m},\boldsymbol{\gamma})\rangle^\alpha}\\
&\leqslant\frac{1}{2^\alpha(2n+1)^\alpha}\left(\sum_{|m_i|\leqslant n^{\frac{\alpha}{n^{\alpha-1}}}}{}'\frac{1}{\|\boldsymbol{m}\|\langle(\boldsymbol{m},\boldsymbol{\gamma})\rangle}\right)^\alpha\\
&\leqslant c(a,b,\alpha,s)n^{(-\alpha+s\alpha(\alpha-1))/(\alpha-1)}(\log 3n)^{\alpha+s\alpha\delta_{1,\alpha}}.
\end{aligned}
\tag{7.2}
$$

显然有

$$\sum\nolimits_2\leqslant\sum_{i=1}^{s}\sum_{|m_i|>n^{\frac{\alpha}{\alpha-1}}}\frac{1}{|m_i|^\alpha}\sum\frac{1}{\|\boldsymbol{m}\|^\alpha}\leqslant c(\alpha,s)n^{-\alpha}.\tag{7.3}$$

将 (7.2)、(7.3) 代入 (7.1) 即得定理 7.1.

显然, 定理 3 是引理 1.1 与引理 7.1 的推论.

## §8. 定理 4 的证明

我们引入记号

$$Q_n(f) = I(f) - \frac{1}{n}\sum_{j=1}^{n} f\left(\frac{j\boldsymbol{\alpha}}{n}\right).$$

**定理 8.1** 若同余式 (6.1) 在区域 (6.2) 中无解, 则

$$\sup_{f\in E_s^\alpha(C)} |Q_n(f)| \leqslant C\cdot c(\alpha,\varepsilon)^s M^{-\alpha+\varepsilon}.$$

**证** 显然可以假定 $\varepsilon<\alpha-1$. 因

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n} f\left(\frac{j\boldsymbol{\alpha}}{n}\right) =&\frac{1}{n}\sum_{j=1}^{n}\sum C(\mathbf{m})\frac{1}{n}\sum_{j=i}^{n} e^{2\pi i(\boldsymbol{\alpha},\boldsymbol{m})j/n}\\ =&C(\mathbf{0})+\sum{}' C(\boldsymbol{m})\frac{1}{n}\sum_{j=1}^{n} e^{2\pi i(\boldsymbol{\alpha},\boldsymbol{m})j/n}\\ =&C(\mathbf{0})+\sum_{(\boldsymbol{\alpha},\boldsymbol{m})\equiv 0(\mathrm{mod} n)}{}' C(\boldsymbol{m}),\end{aligned}$$

由引理 6.2 可知

$$\begin{aligned}&j\sup_{f\in E_s^\alpha(C)} |Q_n(f)| \leqslant C\sum_{(\boldsymbol{\alpha},\boldsymbol{m})\equiv 0(\mathrm{mod} n)}{}'\frac{1}{||\boldsymbol{m}||^\alpha}\\ \leqslant& C\sum_{l=1}^{\infty}\frac{(T_M^{l+1}-T_M^l)}{(lM)^\alpha}\\ =&C\sum_{l=1}^{\infty}T_M^{l+1}\left(\frac{1}{l^\alpha}-\frac{1}{(l+1)^\alpha}\right)\leqslant C\cdot c(\varepsilon)^s M^{-\alpha+\varepsilon}\sum_{l=1}^{\infty}\frac{\alpha}{l^{\alpha-\varepsilon}}\\ \leqslant& C\cdot c(\alpha,\varepsilon)^s M^{-\alpha+\varepsilon},\end{aligned}$$

其中

$$\frac{1}{l^\alpha}-\frac{1}{(l+1)^\alpha}=\alpha\int_l^{l+1}x^{-\alpha-1}dx\leqslant\frac{\alpha}{l^{\alpha+1}}.$$

定理 8.1 证完.

定理 4 可以由引理 1.1, 定理 6.2 及定理 8.1 推出.

## §9. 算　　例

记

$$P_n^*(f) = I(f) - \frac{1}{n}\sum_{j=-(n-1)}^{n-1}\left(1-\frac{|j|}{n}\right) f(j\boldsymbol{\gamma}).$$

首先, 我们证明下面定理:

**定理 9.1**　命 $\gamma_1,\cdots,\gamma_s$ 为一组实数, 而 $1,\gamma_1,\cdots,\gamma_s$ 在有理数域 $R$ 上线性独立. 则

$$\sup_{f\in E_s^2(C)}|P_n^*(f)|\leqslant C\left(\frac{\pi^2}{6}\right)(W(n;\gamma_1,\cdots,\gamma_s)-1),$$

此处

$$W(n;\gamma_1,\cdots,\gamma_s)=\frac{3^s}{n}+\frac{2\cdot 3^s}{n}\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)\prod_{v=1}^{s}(1-2\{\gamma_v j\})^2.$$

**引理 9.1**　我们有

$$\sum_{m=-\infty}^{\infty}\frac{e^{2\pi imx}}{\frac{\pi^2}{6}m^2}=3(1-2\{x\})^2.$$

**证**　因

$$3\int_0^1(1-2x)^2e^{-2\pi imx}dx\begin{cases}0, & 当m=0,\\ \dfrac{6}{\pi^2m^2}, & 当m\neq 0,\end{cases}$$

故得引理 9.1.

**定理 9.1 的证明**　因

$$\sum_{j=-(n-1)}^{n-1}(n-|j|)=\sum_{k=0}^{n-1}\sum_{j=-k}^{k}=n^2$$

及

$$\sum_{j=-(n-1)}^{n-1}(n-|j|)e^{2\pi ij\delta}=\sum_{k=0}^{n-1}\sum_{j=-k}^{k}e^{2\pi ij\delta}$$
$$=\frac{1}{\sin\pi\delta}\sum_{k=0}^{n-1}\sin(2k+1)\pi\delta=\left(\frac{\sin n\pi\delta}{\sin\pi\delta}\right)^2,$$

此处 $\delta$ 为实数, 但非整数, 所以

$$\frac{1}{n}\sum_{j=-(n-1)}^{n-1}\left(1-\frac{|j|}{n}\right)f(j\gamma)$$
$$=\frac{1}{n^2}\sum_{j=-(n-1)}^{n-1}(n-|j|)\sum C(\boldsymbol{m})e^{2\pi i(\boldsymbol{m},\boldsymbol{\gamma})j}$$
$$=\frac{1}{n^2}\sum C(\boldsymbol{m})\sum_{j=-(n-1)}^{n-1}(n-|j|)e^{2\pi i(\boldsymbol{m},\boldsymbol{\gamma})j}$$

$$=C(\mathbf{0})+\frac{1}{n^2}\sum{}'C(\boldsymbol{m})\left(\frac{\sin n\pi(\boldsymbol{m},\boldsymbol{\gamma})}{\sin\pi(\boldsymbol{m},\boldsymbol{\gamma})}\right)^2.$$

由引理 9.1 可知

$$\sup_{f\in E_s^2(C)}|P_n^*(f)|\leqslant\frac{C}{n^2}\sum{}'\frac{1}{||\boldsymbol{m}||^2}\left(\frac{\sin n\pi(\boldsymbol{m},\boldsymbol{\gamma})}{\sin\pi(\boldsymbol{m},\boldsymbol{\gamma})}\right)^2.$$

$$=\frac{C}{n^2}\sum{}'\frac{1}{||\boldsymbol{m}||^2}\sum_{k=0}^{n-1}\sum_{j=-k}^{k}e^{2\pi i(\boldsymbol{m},\boldsymbol{\gamma})j}\leqslant\frac{C}{n^2}\left(\frac{\pi^2}{6}\right)^s\sum_{k=0}^{n-1}\sum_{j=-k}^{k}\sum{}'\frac{e^{2\pi i(\boldsymbol{m},\boldsymbol{\gamma})j}}{\left|\left|\frac{\pi^2}{6}\boldsymbol{m}\right|\right|}$$

$$=\frac{C}{n^2}\left(\frac{\pi^2}{6}\right)^s\sum_{k=0}^{n-1}\sum_{j=-k}^{k}\left(3^s\prod_{v=1}^{s}(1-2\{\gamma_v j\})^2-1\right)$$

$$=C\left(\frac{\pi^2}{6}\right)^s\left(\sum_{j=-(n-1)}^{n-1}(n-|j|)\frac{3^s}{n^2}\prod_{v=1}^{s}(1-2\{\gamma_v j\}^2)-1\right)$$

$$=C\left(\frac{\pi^2}{6}\right)^s(W(n;\gamma_1,\cdots\gamma_s)-1).$$

定理 9.1 证完.

用 Wang520 台式电子计算机计算, 可得下面两表:

**表 1　$s$=3**

| $n$ | $W\left(n;\frac{\sqrt{5}-1}{2},\sqrt{2},\sqrt{10}\right)$ | $W(n;e,e^2,e^3)$ |
|---|---|---|
| 100 | 1.08877 | 1.10689 |
| 500 | 1.01351 | 1.00914 |
| 1000 | 1.00572 | 1.00294 |

**表 2　$s$=4**

| $n$ | $W\left(n;2\cos\frac{2\pi}{11},2\cos\frac{4\pi}{11},2\cos\frac{6\pi}{11},2\cos\frac{8\pi}{11}\right)$ | $W(n;e,e^2,e^3,e^4)$ |
|---|---|---|
| 1000 | 1.03263 | 1.13899 |
| 1500 | 1.02139 | 1.11848 |
| 3000 | 1.00887 | |

## 参 考 文 献

[1] Weyl, H., 1913 *Math. Ann.*, 77, 313～352.

[2] Корбов, Н. М., 1963 Теорстикочисловые методы в приближенном анализе, Издфизмат. лит. Мос.

[3] Roth, K.F., 1954 *Math.*, 1(2), 73~79.
[4] Halton, J.H., 1960 *Num. Math.*, 27(2), 84~ 90.
[5] Шарыгин, Н.Ф., 1963 Журн. мат. ц мат. фцз., 3(2), 370~376.
[6] Бахвалов, Н. С., 1959 Бест. Мос. ун-тα,. 4, 3~18.
[7] Hua Loo Keng and Wang Yuan, 1960 *Scii. Rec,. Acad. Sin.*, 4(1), 8~11.
[8] Hua Loo Keng and Wang Yuan, 1964 On Diophantine Approximations and Numerical Integra tions (I), (II), *Sci. Sin.*, 13(6), 1007~1010.
[9] Hua Loo Keng and Wang Yuan, 1965 *Sci. Sin.*, 14(7), 964~978.
[10] Schmidt, W.M., 1970 *Acta Math.*, 125, 189~201.
[11] Haselgrove, C.B., 1961 *Math. Comp.*, 15(76), 323~337.
[12] Hlawka, E., 1962 *Mon. Math.*, 66(2), 140~151.
[13] Hlawka, E., 1964 *Comp.Math.*, 16(1~2), 92~105.
[14] Baker, A., 1965 *Can. J. Math.*, 17(4), 616~626.
[15] вииоградов, Н. М., 1971 Метод тригонометрических сумм в теории чисел, Рзд. 《Hayka》, Физмат. лит.

**附记** 在本文交稿后, 我们又见到下列两本书 (文献 [16,17]). 一个类似于本文定理 5.1 的定理已由 H.Niederreiter 证明了 [16], 而且文献 [9] 中的某些结果已由 S.Haber 加以改进 [16]. 在文献 [17] 中还总结了某些高维数值积分的数论方法.

[16] Applications of Number Theory to Numerical Analysis, edited by S.K. Zaremba, Acad. Press, 1972.
[17] 津田孝夫, 1973 多变数问题の数值数析, サイコンス株式会社.

# 论一致分布与近似分析 —— 数论方法 (II)*

华罗庚　王　元
(中国科学院数学研究所)

## 提　要

本文将给出前文 [1] 定义的点集贯在数值积分, 插值法与第二类 Fredholm 积分方程近似解问题上的应用. 同时亦研究了由Коробов引入的点集贯.

## §1. 结果的陈述

命 $f(\boldsymbol{x})=f(x_1,\cdots,x_s)$ 为对每一变数都有周期 1 的函数. 命 $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_s)$, 当 $\alpha_k=0$ 时, 命 $\rho_k=\beta_k=0$, 而当 $\alpha_k>0$ 时, 命 $\alpha_k=\rho_k+\beta_k$, 此处 $\rho_k$ 为一非负整数, 而 $0<\beta_k\leqslant 1(1\leqslant k\leqslant s)$. 定义

$$\delta_h^k f=(2i)^{-1}(f(x_1,\cdots,x_k+h,\cdots,x_s)-f(x_1,\cdots,x_k-h,\cdots,x,)).$$

假定诸偏导数

$$\frac{\partial^{\tau_1+\cdots+\tau_s}f}{\partial x_1^{\tau_1}\cdots\partial x_s^{\tau_s}}=f(\boldsymbol{x})^{(\tau_1,\cdots,\tau_s)}\quad(0\leqslant\tau_i\leqslant\rho_i,1\leqslant i\leqslant s)$$

都存在, 而且对每一变数都有周期 1. 定义

$$||f^{\alpha}||=\sup_{\substack{0<hk\leqslant\infty\\ x\in G_s}}\left|\prod_{\alpha_k>0}((h_k^{-\beta_k}\delta_{h_k}^k)f)^{(\rho_1,\cdots,\rho_s)}\right|.$$

命 $H_s^{\alpha}(A)$ 表示满足下面条件

$$||f^{(\theta_1\alpha_1,\cdots,\theta_s\alpha_s)}||\leqslant A$$

的函数构成的函数类, 此处 $\theta_1,\cdots,\theta_s=0$ 或 1 及 $A(>0)$ 为一绝对常数. 特别当 $\alpha_1=\cdots,=\alpha_s=\alpha$ 时, 我们将函数类记为 $H_s^{\alpha}(A)$.

前文 [1] 中引入的记号与规定在此仍将保留. 将前文 [1] 中定义的点集贯 (1.1) 与 (1.2) 式用于数值积分问题, 我们得到如下结果:

* 原载《中国科学》第 17 卷第 3 期, 1947 年 5 月.

**定理 1**　命 $\alpha$ 为适合 $1 \geqslant \alpha > 0$ 的数, 则

$$\sup_{f \in H_q^\alpha(A)} \left| \int_0^1 \cdots \int_0^1 f(x_1, \cdots, x_q) dx_1 \cdots dx_q - \frac{1}{n} \sum_{j=1}^n f(\omega_1 j, \cdots, \omega_q j) \right| \leqslant Ac(\mathscr{R}_r, \alpha, \varepsilon) n^{-\alpha+\varepsilon}. \tag{1.1}$$

**定理 2**　命 $\alpha$ 为适合 $1 \geqslant \alpha > 0$ 的数, 则

$$\sup_{f \in H_r^\alpha(A)} \left| \int_0^1 \cdots \int_0^1 f(x_1, \cdots, \dot{x}_r) dx_1 \cdots dx_r - \frac{1}{n} \sum_{j=1}^n f\left(\frac{j}{n}, \frac{h_1 j}{n}, \cdots, \frac{h_q j}{n}\right) \right| \leqslant Ac(\mathscr{R}_r, \alpha, \varepsilon) n^{-\frac{a}{2} - \frac{a}{2q} + \varepsilon}. \tag{1.2}$$

其次, 我们将用文献 [1] 中定义的点集贯 (1.1) 与 (1.2) 式来处理插值法问题及第二类 Fredholm 型积分方程的渐近解法问题.

我们还研究一致分布点集贯

$$\left(\left\{\frac{j}{p}\right\}, \left\{\frac{j^2}{p}\right\}, \cdots, \left\{\frac{j^s}{p}\right\}\right) \quad (1 \leqslant j \leqslant p) \tag{1.3}$$

与

$$\left(\left\{\frac{j}{p}\right\}, \left\{\frac{aj}{p}\right\}, \cdots, \left\{\frac{a^{s-1}j}{p}\right\}\right) \quad (1 \leqslant j \leqslant p) \tag{1.4}$$

此处 $p$ 为素数, 而 $a = a(p)$ 为依赖于 $p$ 的整数, 这两个点集贯是Коробов[2,3] 分别于 1957 年及 1959 年首先引进的.

**定理 3**　命 $p$ 为素数, 则

$$\sup_{f \in H_s^\alpha(A)} \left| \int_0^1 \cdots \int_0^1 f(x_1 \cdots, x_s) dx_1 \cdots dx_s - \frac{1}{p} \sum_{j=1}^p f\left(\frac{j}{p}, \frac{j^2}{p}, \cdots, \frac{j^s}{p}\right) \right| \leqslant \begin{cases} Ac(\alpha, s) p^{-\frac{1}{2}}, \text{当} 1 \geqslant \alpha > \dfrac{1}{2} \text{时}, \\ Ac(\alpha, s) p^{-\alpha} (\log p)^{s-1+\delta_{\frac{1}{2},\alpha}}, \text{当} \dfrac{1}{2} \geqslant \alpha > 0 \text{时}, \end{cases} \tag{1.5}$$

我们亦建议点集贯

$$\left(\left\{\frac{a}{p}\right\}, \left\{\frac{aj}{p}\right\}, \cdots, \left\{\frac{aj^{s-1}}{p}\right\}\right) \quad (1 \leqslant a, j \leqslant p). \tag{1.6}$$

并将它用于数值积分问题.

研究第二类 Volterra 型积分方程

$$\varphi(x) = \int_0^x K(x, y) \varphi(y) dy + f(x), \tag{1.7}$$

此处 $f(x) \in H_1^\alpha(A)$ 及 $K(x,y) \in H_2^\alpha(A)$. 我们引入记号

$$\nu(\alpha)\begin{cases} \frac{1}{2}, & \text{当}\alpha > 1\text{时}, \\ \frac{2\alpha}{1+4\alpha-\alpha^2}, & \text{当}1 \geqslant \alpha > 0\text{时}, \end{cases} \tag{1.8}$$

$$g = [p^{\nu(\alpha)}], \tag{1.9}$$

$$Q = \left[\nu(\alpha)\alpha\frac{\log_2 p}{\log_2\log_2 p}\right], \tag{1.10}$$

及

$$B_{j,s,p}(a) = \sum_{\bar{m}_1\cdots\bar{m}_s\leqslant g} e^{-2\pi i(m_1+m_2a+\cdots+m_sa^{s-1})j/p} \times \int_0^x\int_0^{x_1}\cdots\int_0^{x_{s-1}} e^{2\pi i(m_1x_1+\cdots+m_sx_s)}dx_1\cdots dx_s. \tag{1.11}$$

**定理 4** 命 $p$ 为素数及 $\alpha$ 为正数, 则存在整数 $a$, 使方程 (1.7) 的解 $\varphi(x)$ 可以写为:

$$\varphi(x) = f(x) + \frac{1}{p}\sum_{j=1}^{p}\sum_{s=1}^{Q}B_{j,s,p}(a)K\left(x,\frac{j}{p}\right)K\left(\frac{j}{p},\frac{aj}{p}\right) \cdots K\left(\frac{a^{s-2}j}{p},\frac{a^{s-1}j}{p}\right)f\left(\frac{a^{s-1}j}{p}\right) + O(p^{-\nu(\alpha)\alpha+\varepsilon}), \tag{1.12}$$

此处与“$O$”有关的常数仅依赖于 $\alpha, A$ 及 $\varepsilon$.

欲得到整数 $a = a(p)$, 我们需要 $c(s)p^2$ 次初等运算. 为了证明定理 3, 我们需要用到 A. Weil[4] 关于指数和估计的著名定理.

**引理 1.1** 命 $p$ 为素数及 $m_1, \cdots, m_s$ 为一组整数, 其中至少有一个不是 $p$ 的倍数, 则

$$\left|\sum_{j=1}^{p} e^{2\pi i(m_1j+\cdots+m_sj^s)/p}\right| \leqslant (s-1)\sqrt{p}.$$

定理 1 及当 $\frac{1}{2} \geqslant \alpha > 0$ 时, 由定理 3 给出的估计, 除因子 $n^\varepsilon$ 及 $(\log p)^{s-1+\delta_{1,\alpha}}$ 外, 是最佳可能的 [1]. 过去在证明 (1.5) 式时, 常需加上较强的限制 $\alpha > \frac{1}{2}$(见文献 [2,5]). 定理 4 则为Шахов[5,6] 结果的改进, 在他原来的结果中, 误差项为 $O(p^{-\frac{\alpha}{2}+\frac{1}{2}+\varepsilon})$, 其中 $\alpha \geqslant 2$ 为整数.

## §2. 函 数 类

命 $c$ 为适合 $0 < c < 1$ 的常数, 命

$$\mu(x) = \begin{cases} \cos^2\left(\frac{\pi}{2}\log_2\left|\frac{x}{c}\right|\right), & \text{当}\frac{c}{2} \leqslant |x| \leqslant 2c\text{时}, \\ 0, & \text{其他情形} \end{cases}$$

及

$$\mu_0(x) = 1 - \sum_{t=1}^{\infty} \mu_t(x), \tag{2.1}$$

此处 $\mu_t(x) = \mu(2^{1-t}x)(t \geqslant 1)$.

命 $f(x)$ 为 $s$- 维对每一变数都有周期 1 的函数, 并有 Fourier 展开式:

$$f(\boldsymbol{x}) \sim \Sigma C(\boldsymbol{m})e^{2\pi i(m,x)}.$$

对于一个以非负整数为分量的矢量 $\boldsymbol{t} = (t_1, \cdots, t_s)$, 定义

$$\varphi_t(\boldsymbol{x}) = \sum C_t(\boldsymbol{m})e^{2\pi i(m,x)},$$

其中

$$C_t(\boldsymbol{m}) = C(\boldsymbol{m})\mu_{t_1}(m_1)\cdots\mu_{t_s}(m_s). \tag{2.2}$$

命 $\boldsymbol{\alpha}$ 为以非负实数为分量的矢量. 满足下面条件的函数 $f(\boldsymbol{x})$ 所构成的函数类记为 $Q_s^{\alpha}(B)$.

$$||\varphi_t|| = \sup_{\boldsymbol{x}\in G_s} |\varphi_t(\boldsymbol{x})| \leqslant B \cdot 2^{-(\boldsymbol{\alpha},\boldsymbol{t})},$$

此处 $B(>0)$ 为绝对常数, 特别当 $\alpha_1 = \cdots = \alpha_s = \alpha$ 时, 将函数类记为 $Q_s^{\alpha}(B)$.

**引理 2.1** 命 $\alpha > 0$, 则

$$H_s^{\alpha}(A) \subset Q_s^{\alpha}(A \cdot C(\alpha)^s) \subset E_s^{\alpha}(A \cdot c(\alpha)^s).$$

**引理 2.2** 命 $\alpha > 0$ 及 $f(\boldsymbol{x}) \in Q_s^{\alpha}(B)$, 则

$$f(\boldsymbol{x}) = \sum{}'' \varphi_t(\boldsymbol{x}),$$

此处 $\Sigma''$ 表示求和号, 其中 $\boldsymbol{t}$ 为过所有有非负整数分量的矢量 [7,8].

## §3. 定理 1 的证明

记

$$R_n(f) = I(f) - \frac{1}{n}\sum_{j=1}^{n} f(j\gamma).$$

显然定理 1 是前文 [1] 引理 1.1, 引理 2.1 及下面定理的推论.

**定理 3.1** 假定 $1 \geqslant \alpha > 0$. 若前文 [1] 中的 (5.1) 式对于任何 $\boldsymbol{m} \neq \boldsymbol{0}$ 皆成立, 则

$$\sup_{f \in Q_s^\alpha(B)} |R_n(f)| \leqslant B \cdot c(a,b,\alpha,s) n^{-\alpha+s(a-1)} (\log 3n)^{s+s\delta_{1,a}}.$$

**证** 由引理 2.2 知, 当 $f \in Q_s^\alpha(B)$ 时, 有

$$R_n(f) = \sum{}'' R_n(\varphi_t). \tag{3.1}$$

故由 $\varphi_t$ 的定义得

$$|R_n(\varphi_t)| \leqslant 2\|\varphi_t\| \leqslant 2B \cdot 2^{-at_0}, \tag{3.2}$$

此处

$$t_0 = t_1 + \cdots + t_s. \tag{3.3}$$

由前文 [1] 的引理 5.1 得

$$\begin{aligned} |R_n(\varphi_t)| &\leqslant \sum{}' + |C_t(\boldsymbol{m})| \frac{1}{n} \left| \sum_{j=1}^n e^{2\pi i (\boldsymbol{m}, \boldsymbol{\gamma}) j} \right| \\ &\leqslant n^{-1} \sum{}' |C_t P(\boldsymbol{m})| \frac{1}{\langle (\boldsymbol{m}, \boldsymbol{\gamma}) \rangle}. \end{aligned} \tag{3.4}$$

因此由 (3.1)~(3.4) 式得

$$\sup_{f \in Q_s^\alpha(B)} |R_n(f)| \leqslant \sup_{f \in Q_s^\alpha(B)} \sum{}'' |R_n(\varphi_t)| \leqslant \sum{}_1 + \sum{}_2, \tag{3.5}$$

其中

$$\sum{}_1 = \sup_{f \in Q_s^\alpha(B)} \sum_{t_0 \leqslant \log_{2n}}{}'' \frac{1}{n} \sum{}' \frac{|C_t(\boldsymbol{m})|}{\langle (\boldsymbol{m}, \boldsymbol{\gamma}) \rangle}$$

及

$$\sum{}_2 = 2B \sum_{t_0 > \log_{2n}} 2^{-at_0}.$$

由 (2.2) 式可知, 当 $\|\boldsymbol{m}\| \geqslant 2^{t_0}$ 时, 有 $C_t(\boldsymbol{m}) = 0$, 所以由前文 [1] 引理 5.5, 引理 2.1 及 (2.1) 式, 得

$$\begin{aligned} \sum{}_1 &\leqslant \sup_{f \in Q_s^\alpha(B)} n^{-1} \sum_{\|\boldsymbol{m}\| \leqslant n}{}' \frac{1}{\langle (\boldsymbol{m}, \boldsymbol{\gamma}) \rangle} \sum{}'' \sum{}'' |C_t(\boldsymbol{m})| \\ &\leqslant \sup_{f \in Q_s^\alpha(B)} n^{-1} \sum_{\|\boldsymbol{m}\| \leqslant n}{}'' \frac{C(\boldsymbol{m})}{\langle (\boldsymbol{m}, \boldsymbol{\gamma}) \rangle} \\ &\leqslant B c(\alpha, s) n^{-1} \sum_{\|\boldsymbol{m}\| \leqslant n}{}' \frac{1}{\|\boldsymbol{m}\|^\alpha \langle (\boldsymbol{m}, \boldsymbol{\gamma}) \rangle} \end{aligned}$$

$$\leqslant Bc(\alpha,s)n^{-1}\sum_{\|\boldsymbol{m}\|\leqslant n}{}'\frac{n^{1-\alpha}}{\|\boldsymbol{m}\|\langle(\boldsymbol{m},\boldsymbol{\gamma})\rangle}$$
$$\leqslant Bc(a,b,\alpha,s)n^{-\alpha+s(\alpha-1)}(\log 3n)^{1+s\delta_{1.\alpha}}. \tag{3.6}$$

因不定方程 (3.3) 的非负整数解 $(t_1,\cdots,t_s)$ 的个数为

$$\begin{pmatrix} t_0+s-1 \\ s-1 \end{pmatrix}=\frac{(t_0+s-1)!}{t_0!(s-1)!}\leqslant c(s)t_0^{s-1},$$

所以

$$\sum\nolimits_2 \leqslant Bc(x)\sum_{t=[\log_2 n]+1}^{\infty}2^{-at}t^{s-1}$$
$$\leqslant Bc(\alpha,s)n^{-\alpha}(\log 3n)^{s-1}.$$

将 (3.6) 与 (3.7) 式代入 (3.5) 式, 即得定理 3.1.

## §4. 定理 2 的证明

我们引入记号

$$S_n(f)=I(f)-\frac{1}{n}\sum_{j=1}^{n}f\left(\frac{j\boldsymbol{\alpha}}{n}\right),$$

此处 $\boldsymbol{\alpha}=(a_1,\cdots,a_s)$ 为有整数分量的矢量.

显然定理 2 是前文 [1] 中引理 1.1、定理 6.2 与下面定理的推论.

**定理 4.1** 命 $\alpha$ 为满足 $1\geqslant\alpha>0$ 的正数, $n\geqslant 2$ 为整数及 $M\geqslant 1$. 若同余式

$$(\alpha,\boldsymbol{m})=\sum_{j=1}^{s}a_i m_i\equiv 0\ (\mathrm{mod} n), \tag{4.1}$$

在区域

$$\|\boldsymbol{m}\|\leqslant M,\boldsymbol{m}\neq\boldsymbol{0} \tag{4.2}$$

中无解, 则

$$\sup_{f\in Q_s^\alpha(B)}|S_n(f)|\leqslant Bc\ (\alpha,\varepsilon)^s M^{-\alpha+\varepsilon}.$$

**证** 显然我们可以假定 $\varepsilon<\alpha$. 类似于 (3.5) 式得

$$\sup_{f\in Q_s^\alpha(B)}|S_n(f)|\leqslant\sum\nolimits_1+\sum\nolimits_2, \tag{4.3}$$

此处

$$\sum\nolimits_1 = \sup_{f\in Q_s^\alpha(B)} \sum_{t_0>\log_2 M}{}'' |S_n(\varphi_t)|$$

与

$$\sum\nolimits_2 = \sup_{f\in Q_s^\alpha(B)} \sum_{t_0\leqslant\log_2 M}{}'' |S_n(\varphi_t)|.$$

由 (3.2) 式可知,

$$\sum\nolimits_1 \leqslant 2B \sum_{t_0>\log_2 M}{}'' 2^{-at_0} \leqslant 2B \sum_{t_0>\log_2 M}{}'' 2^{-(\alpha-\varepsilon)t_0-\varepsilon_0} \leqslant B\cdot c(\alpha,\varepsilon)^s M^{-\alpha+\varepsilon}. \quad (4.4)$$

因当 $\|\boldsymbol{m}\| \geqslant 2^{t_0}$ 时, 有 $C_t(\boldsymbol{m})=0$ 及

$$\begin{aligned}|S_n(\varphi_t)| &= \left|\sum{}' C_t(\boldsymbol{m})\frac{1}{n}\sum_{j=1}^{n} e^{2\pi i(\alpha,\boldsymbol{m})j/n}\right| \\ &\leqslant \sum_{(\boldsymbol{\alpha},\boldsymbol{m})\equiv 0(\mathrm{mod} n)}{}' |C_t(\boldsymbol{m})|,\end{aligned}$$

所以

$$\sum\nolimits_2 \leqslant \sup_{f\in Q_s^\alpha(B)} \sum_{(\boldsymbol{\alpha},\boldsymbol{m})\equiv 0(\mathrm{mod} n)}{}' \sum_{t_0\leqslant\log_2 M}{}'' |C_t(\boldsymbol{m})| = 0. \quad (4.5)$$

故由 (4.3)~(4.5) 式即得定理 4.1.

## §5. 定理 3 的证明

**定理 3 的证明** 记

$$T_p(f) = I(f) - \frac{1}{p}\sum_{j=1}^{p} f\left(\frac{j}{p},\frac{j^2}{p},\cdots,\frac{j^s}{p}\right).$$

由引理 2.1 与引理 2.2, 可知

$$\sup_{f\in H_s^\alpha(A)} |T_p(f)| \leqslant \sum\nolimits_1 + \sum\nolimits_2 \quad (5.1)$$

此处

$$\sum\nolimits_1 = \sup_{f\in H_s^\alpha(A)} \sum_{t_0\geqslant\log_2 \mathrm{p}}{}'' |T_p(\varphi_t)|$$

与

$$\sum\nolimits_2 = \sup_{f\in H_s^\alpha(A)} \sum_{t_0<\log_2 \mathrm{p}}{}'' |T_p(\varphi_t)|$$

类似于 (3.7) 式, 得

$$\begin{aligned}\sum\nolimits_1 &\leqslant 2\sup_{f\in H_s^\alpha(A)}\sum_{t_0\geqslant\log_2 p}{}''\|\varphi_t\|\\ &\leqslant Ac(\alpha,s)\sum_{t=[\log_2 p]}^{\infty}2^{-\alpha t}t^{s-1}\\ &\leqslant Ac(\alpha,s)p^{-\alpha}(\log p)^{s-1}\end{aligned}\tag{5.2}$$

命 $\boldsymbol{m}\neq 0$. 若诸关系 $|m_i|<2^{t_i}(1\leqslant i\leqslant s)$ 中有一个不满足, 则 $C_t(m)=0$. 若对于 $1\leqslant i\leqslant s$ 中所有的 $i$ 皆有 $|m_i|<2^{t_i}$, 此处 $t_0<\log_2 p$, 则 $p$ 不能整除所有的 $m_i$, 故由引理 1.1 得

$$\begin{aligned}\sum\nolimits_2 &\leqslant \sup_{f\in H_s^\alpha(A)}\sum_{t_0<\log_2 p}{}''\sum{}'|C_t(m)|\left|\frac{1}{p}\sum_{j=1}^{p}e^{2\pi i(m_1 j+\cdots+m_s j^s)/p}\right|\\ &\leqslant(s-1)p^{-\frac{1}{2}}\sup_{f\in H_s^\alpha(A)}\sum_{t_0<\log_2 p}{}''\sum_{|m_i|<2^{t_i}}{}'|C_t(m)|.\end{aligned}$$

从而由 Schwarz 不等式及

$$\|\varphi_t\|_{L_2}\leqslant\|\varphi_t\|\leqslant Ac(\alpha,s)a^{-at_0},$$

此处

$$\|\varphi_t\|_{L_2}=\left(\int_{G_s}|\varphi_t|^2\mathrm{d}\boldsymbol{x}\right)^{\frac{1}{2}},$$

我们得

$$\begin{aligned}\sum\nolimits_2 &\leqslant(s-1)p^{-\frac{1}{2}}\sup_{f\in H_s^\alpha(A)}\sum_{t_0<\log_2 p}{}''\Big(\sum_{|m_i|<2^{t_i}}1\Big)^{\frac{1}{2}}\Big(\sum_{|m_i|<2^{t_i}}|C_t(\boldsymbol{m})|^2\Big)^{\frac{1}{2}}\\ &\leqslant(s-1)p^{\frac{1}{2}}\sup_{f\in H_s^\alpha(A)}\sum_{t_0<\log_2 p}{}''2^{\frac{s+t_0}{2}}\|\varphi_t\|_{L_2}\\ &\leqslant Ac(\alpha,s)p^{-\frac{1}{2}}\sum_{t_0<\log_2 p}{}''2^{-(\alpha-\frac{1}{2})t_0}\\ &\leqslant Ac(\alpha,s)p^{-\frac{1}{2}}\sum_{t=0}^{[\log_2 p]}2^{-(\alpha-\frac{1}{2})t}(t+1)^{s-1}\\ &\leqslant\begin{cases}Ac(\alpha,s)p^{-\frac{1}{2}}, & \text{当}1\geqslant\alpha>\dfrac{1}{2}\text{时},\\ Ac(\alpha,s)p^{-a}(\log p)^{s-1+\delta_{\frac{1}{2},\alpha}}, & \text{当}\dfrac{1}{2}\geqslant\alpha>0\text{时}.\end{cases}\end{aligned}\tag{5.3}$$

将 (5.2)、(5.3) 式代入 (5.1) 式即得定理 4.1.

**附记** 对于 $\alpha > 1$, 我们建议用一致分布点集贯 (1.6) 式, 则得下面的结果: 命 $p$ 为素数及 $n = p^2$, 则

$$\sup_{f\in E_s^\alpha(C)}\left|I(f)-\frac{1}{n}\sum_{a=1}^{p}\sum_{j=1}^{p}f\left(\frac{a}{p},\frac{aj}{p},\cdots,\frac{aj^{s-1}}{p}\right)\right|$$
$$\leqslant C\cdot c(\alpha,s)n^{-\frac{1}{2}}.$$

**证** 当 $f\in E_s^\alpha(C)$ 时, 我们显然有

$$\frac{1}{n}\sum_{a=1}^{p}\sum_{j=1}^{p}f\left(\frac{a}{p},\frac{aj}{p},\cdots,\frac{aj^{s-1}}{p}\right)$$
$$=\frac{1}{p^2}\sum C(\boldsymbol{m})\sum_{a=1}^{p}\sum_{j=1}^{p}e^{2\pi i(m_1+m_2j+\cdots+m_sj^{s-1})a/p}$$
$$=I(f)+p^{-1}{\sum}'C(\boldsymbol{m})\sum_{\substack{m_1+\cdots+m_sj^{s-1}\equiv 0(\mathrm{mod}p)\\ 1\leqslant j\leqslant p}}1.$$

若诸 $m_i$ 中至少有一个非 $p$ 的倍数, 则同余式

$$m_1+\cdots+m_sj^{s-1}\equiv 0(\mathrm{mod}p)(1\leqslant j\leqslant p)$$

的解的个数最多为 $s-1$, 所以

$$\sup_{f\in E_s^\alpha(C)}\left|I(f)-\frac{1}{n}\sum_{a=1}^{p}\sum_{j=1}^{p}f\left(\frac{a}{p},\frac{aj}{p},\cdots,\frac{aj^{s-1}}{p}\right)\right|$$
$$\leqslant Cp^{-1}{\sum}'\frac{1}{||\boldsymbol{m}||^\alpha}\sum_{\substack{m_1+\cdots+m_sj^{s-1}\equiv 0(\mathrm{mod}p)\\ 1\leqslant j\leqslant p}}1$$
$$\leqslant C\cdot(s-1)p^{-1}\sum\frac{1}{||\boldsymbol{m}||^\alpha}+C{\sum}'\frac{1}{||p\boldsymbol{m}||^\alpha}$$
$$\leqslant C\cdots 2sp^{-1}\sum\frac{1}{||\boldsymbol{m}||^\alpha}$$
$$\leqslant C\cdot c(\alpha,s)n^{-\frac{1}{2}}.$$

附记证完.

## §6. 插　值　法

我们引入记号

$$\varDelta=\sup_{f\in Q_s^\alpha(B)}||f(\boldsymbol{x})-\frac{1}{n}\sum_{j=1}^{n}f\left(\frac{j\boldsymbol{\alpha}}{n}\right)\sum_{||\boldsymbol{m}||\leqslant n_1}e^{2\pi i(\boldsymbol{m},\boldsymbol{x}-\frac{j\boldsymbol{\alpha}}{n})}||_{L_2}.$$

**定理 6.1**　若同余式 (4.1) 在区域 (4.2) 式中无解, 则

$$\Delta \leqslant Bs!^{\frac{1}{2}} c(\alpha,\varepsilon)^s M^{-\nu(\alpha)\alpha+\varepsilon},$$

此处

$$n_1 = [M^{\nu(\alpha)}],$$

其中 $\nu(\alpha)$ 由 (1.8) 式定义

**引理 6.1**　命 $\boldsymbol{l} = (l_1, \cdots, l_s)$ 为以整数为分量的矢量, 并且满足 $||\boldsymbol{l}|| \geqslant 3^s$, 又假定 $N$ 满足 $1 \leqslant N \leqslant ||\boldsymbol{l}||/3s$, 则

$$\sum_{||\boldsymbol{m}|| \leqslant N} \frac{1}{||\boldsymbol{l}+\boldsymbol{m}||^\alpha} < \begin{cases} s!c(\alpha,\varepsilon)^s N^{1+\varepsilon}/||\boldsymbol{l}||^\alpha, & \text{当} 1 \geqslant \alpha > 0 \text{时}, \\ s!c(\alpha)^s N^\alpha/||\boldsymbol{l}||^\alpha, & \text{当} \alpha > 1 \text{时}. \end{cases}$$

**证**　首先假定 $1 \geqslant \alpha > 0$. 因当 $s=1$ 时, $N \leqslant \bar{l}_1/3$, 所以

$$\sum_{\bar{m}_1 \leqslant N} \frac{1}{(\overline{l_1+m_1})^\alpha} \leqslant \left(\frac{3}{2}\right)^\alpha \frac{3N}{\bar{l}_1^\alpha},$$

所以引理对于 $s=1$ 成立. 设 $k$ 为正整数及引理 6.1 对于 $s=1,2\cdots,k$ 时成立, 现证明引理对于 $s=k+1$ 时亦成立.

显然, 由 $\bar{m}_1\cdots\bar{m}_{k+1} \leqslant N \leqslant \bar{l}_1\cdots\bar{l}_{k+1}/3^{k+1}$ 可知, 至少存在一个 $m_i$, 使 $\bar{m}_i < \bar{l}_i/2$, 此处 $1 \leqslant i \leqslant k+1$, 所以

$$\sum_{\bar{m}_1\cdots\bar{m}_{k+1} \leqslant N} \frac{1}{((\overline{l_1+m_1})\cdots(\overline{l_{k+1}+m_{k+1}}))^\alpha} \leqslant \sum\nolimits_1 + \cdots + \sum\nolimits_{k+1},$$

此处

$$\sum\nolimits_i = \sum_{\substack{\bar{m}_1\cdots\bar{m}_{k+1} \leqslant N \\ \bar{m}_i < \bar{l}_i/2}} \frac{1}{((\overline{l_1+m_1})\cdots(\overline{l_{k+1}+m_{k+1}}))^\alpha} (1 \leqslant i \leqslant k+1).$$

(1) 假定 $N \leqslant \bar{l}_2\cdots\bar{l}_{k+1}/3^k$, 则由归纳法假定可得

$$\begin{aligned} \sum\nolimits_1 &\leqslant \sum_{\bar{m}_1 \leqslant N} \frac{2^\alpha}{\bar{l}_1^\alpha} \sum_{\bar{m}_2\cdots\bar{m}_{k+1} \leqslant N/\bar{m}_1} \frac{1}{((\overline{l_2+m_2})\cdots(\overline{l_{k+1}+m_{k+1}}))\alpha} \\ &\leqslant \frac{k!c(\alpha,\varepsilon)^k N^{1+\varepsilon}}{(\bar{l}\cdots\bar{l}_{k+1})^\alpha} \sum_{\bar{m}_1 \leqslant N} \frac{1}{(\bar{m}_1)^{1+\varepsilon}} \\ &\leqslant \frac{k!c(\alpha,\varepsilon)^{k+1} N^{1+\varepsilon}}{(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha}. \end{aligned}$$

(2) 假定 $N > \bar{l}_2\cdots\bar{l}_{k+1}/3^k$, 则

$$\sum\nolimits_2 \leqslant \sigma_1 + \sigma_2,$$

此处

$$\sigma_1=\frac{2^\alpha}{\bar{l}_1^\alpha}\sum_{\bar{m}_1\leqslant 3^kN/\bar{l}_2\cdots\bar{l}_{k+1}}\sum_{\bar{m}_2\cdots\bar{m}_{k+1}\leqslant N/\bar{m}_1}\frac{1}{((\overline{l_2+m_2})\cdots(\overline{l_{l+1}+m_{k+1}}))^\alpha}$$

及

$$\sigma_2=\frac{2^\alpha}{\bar{l}_1^\alpha}\sum_{3^kN/\bar{l}_2\cdots\bar{l}_{k+1}<\bar{m}_1\leqslant\bar{l}_1/2}\sum_{\bar{m}_2\cdots\bar{m}_{k+1}\leqslant N/\bar{m}_1}\frac{1}{((\overline{l_2+m_2})\cdots(\overline{l_{k+1}+m_{k+1}}))^\alpha}.$$

显然

$$\begin{aligned}\sigma_1\leqslant&\frac{2^\alpha}{\bar{l}_1^\alpha}\sum_{\bar{m}_1\leqslant 3^kN/\bar{l}_2\cdots\bar{l}_{k+1}}\sum_{\bar{m}_2\cdots\bar{m}_{k+1}\leqslant N/\bar{m}_1}\frac{(\bar{m}_2\cdots\bar{m}_{k+1})^{1-\alpha+\varepsilon}}{(\bar{m}_2\cdots\bar{m}_{k+1})^{1+\varepsilon}}\\ \leqslant&\frac{c(\alpha,\varepsilon)^kN^{1-\alpha+\varepsilon}}{\bar{l}_1^\alpha}\sum_{\bar{m}_1\leqslant 3^kN/\bar{l}_2\cdots\bar{l}_{k+1}}\frac{(\bar{m}_1)^\alpha}{(\bar{m}_1)^{1+\varepsilon}}\\ \leqslant&c(\alpha,\varepsilon)^k\frac{N^{1+\varepsilon}}{(\bar{l}_1\cdots\bar{l}_{k+1}^\alpha)}\sum_{m_1=-\infty}^{\infty}\frac{1}{(\bar{m}_1)^{1+\varepsilon}}\\ \leqslant&c(\alpha,\varepsilon)^{k+1}N^{1+\varepsilon}/(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha.\end{aligned}$$

又由归纳法假定得

$$\begin{aligned}\sigma_2\leqslant&\frac{k!c(\alpha,\varepsilon)^kN^{1+\varepsilon}}{(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha}\sum_{m_1=-\infty}^{\infty}\frac{1}{(\bar{m}_1)^{1+\varepsilon}}\\ \leqslant&\frac{k!c(\alpha,\varepsilon)^{k+1}N^{1+\varepsilon}}{(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha}.\end{aligned}$$

(3) 由 (1)、(2) 可得

$$\sum\nolimits_1\leqslant\frac{k!c(\alpha,\varepsilon)^{k+1}N^{1+\varepsilon}}{(\bar{l}_1\cdots\bar{l}_{k+1})^\alpha}.$$

因诸 $\Sigma_i$ 适合同样的关系式, 故由归纳法可知, 当 $1\geqslant\alpha>0$ 时, 引理 6.1 成立.

我们可以类似地处理 $\alpha>1$ 的情况.

引理 6.1 证完

**附记** 当 $1\geqslant\alpha>0$ 时, 上述引理的结论可以改为

$$\sum_{||\boldsymbol{m}||\leqslant N}\frac{1}{||\boldsymbol{l}+\boldsymbol{m}||^\alpha}<s!c(\alpha)^s\frac{N(\log 3N)^{s-1}}{||\boldsymbol{l}||^\alpha}.$$

**引理 6.2** 假定 $Q\geqslant 1$ 及 $1\geqslant\alpha>0$, 若同余式 (4.1) 在区域 (4.2) 式中无解, 则

$$\sum_{\substack{(\boldsymbol{a},\boldsymbol{m})\equiv 0(\mathrm{mod}\,n)\\ ||\boldsymbol{m}||\leqslant Q}}\frac{1}{||\boldsymbol{m}||^\alpha}\leqslant c(\varepsilon)^sQ^{1-\alpha+\varepsilon}M^{-1}.$$

**证**　由文献 [1] 的定理 8.1, 得

$$\sum_{(\boldsymbol{a},\boldsymbol{m})\equiv 0(\bmod n)}{}' \frac{1}{||\boldsymbol{m}||^{1+\varepsilon}} \leqslant c(\varepsilon)^s M^{-1},$$

所以

$$\sum_{\substack{(\boldsymbol{a},\boldsymbol{m})\equiv 0(\bmod n)\\ ||\boldsymbol{m}||\leqslant Q}}{}' \frac{1}{||\boldsymbol{m}||^{\alpha}} \leqslant Q^{1-\alpha+\varepsilon} \sum_{(\boldsymbol{a},\boldsymbol{m})\equiv 0(\bmod n)}{}' \frac{1}{||\boldsymbol{m}||^{1+\varepsilon}} \leqslant c(\varepsilon)^s Q^{1-\alpha+\varepsilon} M^{-1}.$$

引理 6.2 证毕.

**定理 6.1 的证明**　我们显然可以假定 $\varepsilon < \alpha\nu(\alpha)$, 取

$$T = \begin{cases} [\log_2 M] + 1, & \text{当}\alpha > 1\text{时}, \\ [\log_2 M n_1^{-1+\frac{\alpha}{2}}] + 1, & \text{当}1 \geqslant \alpha > 0\text{时}. \end{cases}$$

由 Minkowski 不等式, 得

$$\varDelta \leqslant \varDelta_1 + \varDelta_2 + \varDelta_3, \tag{6.1}$$

其中

$$\varDelta_1 = \sup_{f\in Q_s^\alpha(B)} ||f(\boldsymbol{x}) - \sum_{t_0\leqslant T}{}'' \varphi_t(\boldsymbol{x})||_{L_2},$$

$$\varDelta_2 = \sup_{f\in Q_s^\alpha(B)} ||\sum_{t_0\leqslant T}{}'' \left(\varphi_t(\boldsymbol{x}) - \frac{1}{n}\sum_{j=1}^{n} \varphi_t\left(\frac{j\boldsymbol{\alpha}}{n}\right) \sum_{||\boldsymbol{m}||\leqslant n_1} e^{2\pi i(m,\boldsymbol{x}-\frac{j\boldsymbol{\alpha}}{n})}\right) ||_{L_2},$$

$$\varDelta_3 = \sup_{f\in Q_s^\alpha(B)} ||\frac{1}{n}\sum_{j=1}^{n} f\left(\frac{j\boldsymbol{\alpha}}{n}\right) \sum_{||m||\leqslant n_1} e^{2\pi i(m,\boldsymbol{x}-\frac{j\boldsymbol{\alpha}}{n})}$$
$$- \sum_{t_0\leqslant T}{}'' \frac{1}{n}\sum_{j=1}^{n} \varphi_t\left(\frac{j\boldsymbol{\alpha}}{n}\right) \sum_{||m||\leqslant n_1} e^{2\pi i(m,x-\frac{j\alpha}{n})}||_{L_2}.$$

(1) 由 Minkowski 不等式及引理 2.2, 可知

$$\begin{aligned} \varDelta_1 &\leqslant \sup_{f\in Q_s^\alpha(B)} \sum_{t_0>T}{}'' ||\varphi_t||_{L_2} \\ &\leqslant B \sum_{t_0>T}{}'' 2^{-(\alpha-\varepsilon)t_0-\varepsilon t_0} \\ &\leqslant B c(\alpha,\varepsilon)^s 2^{-(\alpha-\varepsilon)T} \\ &\leqslant B c(\alpha,\varepsilon)^s M^{-\nu(\alpha)\alpha+\varepsilon}. \end{aligned} \tag{6.2}$$

(2) $\Delta_2 \leqslant \sigma_1 + \sigma_2$, 此处

$$\sigma_1 = \sup_{f \in Q_s^\alpha(B)} \|\sum_{t_0 \leqslant T}'' \sum_{\|\boldsymbol{m}\| \leqslant n_1} \left( C_t(\boldsymbol{m}) - \frac{1}{n} \sum_{j=1}^{n} \varphi_t \left( \frac{j\boldsymbol{a}}{n} \right) e^{-2\pi i(\boldsymbol{m}, \frac{j\boldsymbol{a}}{n})} \right) e^{2\pi i(\boldsymbol{m},\boldsymbol{x})} \|_{L_2},$$

$$\sigma_2 = \sup_{f \in Q_s^\alpha(B)} \|\sum_{t_0 \leqslant T} \sum_{\|m\| > n_1} C_t(\boldsymbol{m}) e^{2\pi i(\boldsymbol{m},\boldsymbol{x})} \|_{L_2}.$$

因

$$C_t(\boldsymbol{m}) - \frac{1}{n} \sum_{j=1}^{n} \varphi_t \left( \frac{j\boldsymbol{\alpha}}{n} \right) e^{-2\pi i(\boldsymbol{m}, \frac{j\boldsymbol{\alpha}}{n})}$$
$$= - \sum_{(\alpha,l) \equiv 0 (\mathrm{mod} n)}' C_t(\boldsymbol{l} + \boldsymbol{m})$$

及

$$\|\boldsymbol{l}\| \leqslant 2^s \|\boldsymbol{m}\| \cdot \|\boldsymbol{l} + \boldsymbol{m}\|,$$

所以

$$\begin{aligned}
\sigma_1^2 &\leqslant \sup_{f \in Q_s^\alpha(B)} \left\| \sum_{t_0 \leqslant T}'' \sum_{\|\boldsymbol{m} \leqslant n_1} \sum_{(\boldsymbol{\alpha},\boldsymbol{l}) \equiv 0 (\mathrm{mod} n)}' C_t(\boldsymbol{l} + \boldsymbol{m}) e^{2\pi i(\boldsymbol{m},\boldsymbol{x})} \right\|_{L_2}^2 \\
&= \sup_{f \in Q_s^\alpha(B)} \sum_{\|\boldsymbol{m} \leqslant n_1} \Big( \sum_{t_0 \leqslant T}'' \sum_{(\boldsymbol{\alpha},\boldsymbol{l}) \equiv 0 (\mathrm{mod} n)}' |C_t(\boldsymbol{l} + \boldsymbol{m})| \Big)^2 \\
&\leqslant B^2 c(\alpha)^s \sum_{\|\boldsymbol{m}\| \leqslant n_1} \Big( \sum_{\substack{(\boldsymbol{\alpha},\boldsymbol{l}) \equiv 0 (\mathrm{mod} n) \\ \|\boldsymbol{l}\| \leqslant 2^s + T n_1}}' \frac{1}{\|\boldsymbol{l} + \boldsymbol{m}\|^\alpha} \Big)^2 \\
&\leqslant B^2 c(\alpha)^s n_1^\alpha \sum_{\|\boldsymbol{m}\| \leqslant n_1} \sum_{\substack{(\boldsymbol{\alpha},\boldsymbol{l}) \equiv 0 (\mathrm{mod} n) \\ \|\boldsymbol{l}\| \leqslant 2^s + T n_1}}' \frac{1}{\|\boldsymbol{l} + \boldsymbol{m}\|^\alpha} \sum_{\substack{(\boldsymbol{\alpha},\boldsymbol{l}') \equiv 0 (\mathrm{mod} n) \\ \|\boldsymbol{l}'\| \leqslant 2^s + T n_1}}' \frac{1}{\|\boldsymbol{l}'\|^\alpha}.
\end{aligned}$$

显然, 我们可以假定 $n_1 \leqslant M/3^s$, 因此由文献 [1] 中定理 8.1、引理 6.1 与引理 6.2, 得

$$\sigma_1^2 \leqslant \begin{cases} B^2 s! c(\alpha, \varepsilon)^s n_1^{2\alpha} M^{-2\alpha + 2\varepsilon}, & \text{当}\alpha > 1\text{时}, \\ B^2 s! c(\alpha, \varepsilon)^s n_1^{3-\alpha+\varepsilon} M^{-2} 2^{2T(1-\alpha)+T\varepsilon}, & \text{当}1 \geqslant \alpha > 0\text{时} \end{cases}$$

由 Schwarz 不等式得

$$\begin{aligned}
\sigma_2^2 &= \sup_{f \in Q_s^\alpha(B)} \sum_{\|\boldsymbol{m}\| > n_1} \Big( \sum_{t_0 \leqslant T}'' |C_t(\boldsymbol{m})| \Big)^2 \\
&\leqslant \sup_{f \in Q_s^\alpha(B)} \sum_{\|\boldsymbol{m}\| > n_1} \sum_{t_0 \leqslant T}'' 2^{-\frac{\varepsilon t_0}{2}} \sum_{t_0 \leqslant T}'' 2^{\frac{\varepsilon t_0}{2}} |C_t(\boldsymbol{m})|^2
\end{aligned}$$

$$
\begin{aligned}
&\leqslant c(\varepsilon)^s \sup_{f\in Q_s^{\alpha}(B)} \sum_{t_0\leqslant T}{}'' \sum_{\|\boldsymbol{m}\|>n_1} 2^{\frac{\varepsilon t_0}{2}} |C_t(\boldsymbol{m})|^2 \\
&\leqslant c(\varepsilon)^s \sup_{f\in Q_s^{\alpha}(B)} \sum_{t_0>\log_2 n_1}{}'' 2^{\frac{\varepsilon t_0}{2}} \|\varphi_t\|_{L_2}^2 \\
&\leqslant B^2 c(\varepsilon)^s \sum_{t_0>\log_2 n_1}{}'' 2^{-2\alpha t_0+\frac{\varepsilon t_0}{2}} \\
&\leqslant B^2 c(a,\varepsilon)^s n_1^{-2\alpha+\varepsilon}.
\end{aligned}
$$

所以

$$
\Delta_2 \leqslant Bs!c(\alpha,\varepsilon)^s M^{-\nu(\alpha)\alpha+\varepsilon} \tag{6.3}
$$

$$
\begin{aligned}
(3)\ \Delta_3 &= \sup_{f\in Q_s^{\alpha}(B)} \Big\| \sum_{t_0>T}{}'' \frac{1}{n}\sum_{j=1}^{n} \varphi_t\left(\frac{j\boldsymbol{\alpha}}{n}\right) \sum_{\|\boldsymbol{m}\|\leqslant n_1} e^{2\pi i(\boldsymbol{m},\boldsymbol{x}-\frac{j\boldsymbol{\alpha}}{n})} \Big\|_{L_2} \\
&\leqslant \sum_{t_0>T}{}'' \sup_{f\in Q_s^{\alpha}(B)} \|\varphi_t\| \Big( \sum_{\|\boldsymbol{m}\|\leqslant n_1} 1 \Big)^{\frac{1}{2}} \\
&\leqslant B \sum_{t_0>T}{}'' 2^{-at_0} \Big( \sum_{\|\boldsymbol{m}\|\leqslant n_1} \frac{n_1^{1+\varepsilon}}{\|\boldsymbol{m}\|^{1+\varepsilon}} \Big)^{\frac{1}{2}} \\
&\leqslant Bc(\alpha,\varepsilon) n_1^{\frac{1}{2}+\frac{\varepsilon}{2}} 2^{-\alpha T+\frac{\varepsilon T}{2}} \\
&\leqslant Bc(\alpha,\varepsilon)^s M^{-\nu(\alpha)\alpha+\varepsilon}.
\end{aligned} \tag{6.4}
$$

故由 (6.1)~(6.4) 式即得定理 6.1.

**附记** 上面定理的结论可以换为

$$
\Delta \leqslant Bc(\alpha,s) M^{-\nu(\alpha)\alpha} (\log 3M)^{\nu_1^{(\alpha)}},
$$

此处

$$
\nu_1(\alpha) = \begin{cases} s-1, \text{当}\alpha>1\text{时}, \\ \dfrac{(1-5\alpha+\alpha^2)(s-1)+2\alpha^2\delta_{1,\alpha}}{1+4\alpha-\alpha^2}, \text{当}1\geqslant\alpha>0\text{时}. \end{cases}
$$

现在我们将文献 [1] 定义的点集贯 (1.2) 式用于插值法问题. 引入记号 $\boldsymbol{h}=(1,h_1,\cdots,h_q)$, 此处 $(h_1,\cdots,h_q;n)$ 为由文献 [1] 中 (1.2) 式所定义的整数集合, 则得

**定理 6.2** 我们有

$$
\begin{aligned}
\Delta &= \sup_{f\in H_r^{\alpha}(A)} \Big\| f(\boldsymbol{x}) - \frac{1}{n}\sum_{j=1}^{n} f\left(\frac{j\boldsymbol{h}}{n}\right) \sum_{\|\boldsymbol{m}\|\leqslant n_1} e^{-2\pi i(\boldsymbol{m},\frac{j\boldsymbol{h}}{n})} e^{2\pi i(\boldsymbol{m},\boldsymbol{x})} \Big\|_{L_2} \\
&\leqslant Ac(\mathscr{R}_r,\varepsilon) n^{-\nu(\alpha)\alpha(q+1)/2q+\varepsilon},
\end{aligned}
$$

此处

$$n_1 = \left[n^{\frac{\nu(\alpha)(q+1)}{2q}}\right].$$

## §7. 插 值 法 (续)

**定理 7.1** 命 $p$ 为素数及 $n_1 = [p^{\nu(\alpha)}]$, 则存在矢量 $\boldsymbol{a} = (1, a, \cdots, a^{s-1})$(其中 $a$ 为整数), 使

$$\sup_{f \in Q_s^\alpha(B)} ||f(\boldsymbol{m}) - \frac{1}{p}\sum_{j=1}^{p} f\left(\frac{j\boldsymbol{a}}{n}\right) \sum_{||\boldsymbol{m}|| \leqslant n_1} e^{-2\pi i(\boldsymbol{m}, \frac{j\boldsymbol{a}}{n})} e^{2\pi i(\boldsymbol{m}, \boldsymbol{x})}||_{L_2}$$
$$\leqslant Bs!^{\frac{1}{2}} c(\alpha, \varepsilon)^s p^{-\nu(\alpha)\alpha + \varepsilon}.$$

**证** 由定理 6.1 可知, 只要证明存在整数 $a$, 使同余式

$$(\boldsymbol{a}, \boldsymbol{m}) = m_1 + m_2 a + \cdots + m_s a^{s-1} \equiv 0 (\bmod p) \tag{7.1}$$

在区域

$$||\boldsymbol{m}|| \leqslant c(s, \varepsilon) p^{1-\varepsilon}, \boldsymbol{m} \neq \boldsymbol{0} \tag{7.2}$$

中无解即可, 此处

$$c(s, \varepsilon) = \left(2 \sum_{m=-\infty}^{\infty} \frac{1}{\bar{m}^{1+\varepsilon}}\right)^{-s}.$$

显然, 我们可以假定 $c(s, \varepsilon) p^{1-\varepsilon} \geqslant 1$. 取区域 (7.2) 式的一个矢量, 则同余式 (7.1) 在区间 $1 \leqslant \alpha \leqslant p$ 中的解数不超过 $s-1$, 所以, 同余式 (7.1) 在区域 (7.2) 式与 $1 \leqslant a \leqslant p$ 中的解数为:

$$\sum_{||\boldsymbol{m}|| \leqslant c(s,\varepsilon)p^{1-\varepsilon}}{}' \sum_{\substack{(\boldsymbol{a}, \boldsymbol{m}) \equiv 0 (\bmod n) \\ 1 \leqslant a \leqslant p}} 1 \leqslant (s-1) \sum_{||\boldsymbol{m}|| \leqslant c(s,\varepsilon)p^{1-\varepsilon}}{}' \frac{||\boldsymbol{m}||^{1+\varepsilon}}{||\boldsymbol{m}||^{/1+\varepsilon}}$$
$$\leqslant (s-1) c(s, \varepsilon) p \left(\sum_{m=-\infty}^{\infty} \frac{1}{(\bar{m})^{1+\varepsilon}}\right)^s < \frac{p}{2}.$$

所以, 存在整数 $a$ 满足 $1 \leqslant a \leqslant p$, 使同余式 (7.1) 在区域 (7.2) 式中无解.

定理 7.1 证完

**定理 7.2** 命 $n \geqslant 3$ 为整数, 则对于任何满足 $n > n_1 \geqslant 1$ 的整数 $n_1$ 及以整数为分量的矢量 $\boldsymbol{a} = (a_1, \cdots, a_s)$ 皆有

$$\varDelta = \sup_{f \in H_s^\alpha(A)} ||f(\boldsymbol{x}) - \frac{1}{n}\sum_{j=1}^{n} f\left(\frac{j\boldsymbol{a}}{n}\right) \sum_{||\boldsymbol{m}|| \leqslant n_1} e^{-2\pi i(\boldsymbol{m}, \frac{j\boldsymbol{a}}{n})} e^{2\pi i(\boldsymbol{m}, \boldsymbol{x})}||_{L_2}$$

$$\geqslant \frac{A}{(4\pi)^{\alpha+1}} n^{-\frac{\alpha}{2}}.$$

**证** 我们显然可以假定 $a_1=-1$ 及 $(a_2,n)=1$. 命 $\frac{p_t}{q_t}$ 为 $\frac{p_h}{q_h}=\frac{a_2}{n}$ 的 $t$ 次渐近分数, 假定 $n_1$ 满足

$$1=q_0<\cdots<q_k\leqslant n_1<q_{k+1}\leqslant\cdots\leqslant q_h\leqslant n.$$

命 $K=(p_hq_k-q_hp_k)$, 则

$$|K|\leqslant q_h/q_{k+1}\leqslant n/n_1.$$

取 $H_s^\alpha(A)$ 的函数:

$$f(\boldsymbol{x})=\frac{A}{4(2\pi)^{\alpha+1}}\left(\frac{e^{2\pi i(kx_1+x_2)}}{|K|^\alpha}+\frac{e^{-2\pi i(kx_1+x_2)}}{|K|^\alpha}+\frac{e^{2\pi i(n_1+1)x_1}}{(n_1+1)^\alpha}\frac{e^{-2\pi i(n_1+1)x_1}}{(n_1+1)^\alpha}\right).$$

则得

$$\begin{aligned}\Delta^2 &\geqslant \sum_{\bar m_1\bar m_2\leqslant n_1}\Big|\sum_{l_1\equiv a_2l_2(\mathrm{mod} n)}{}'C(l_1+m_1,l_2+m_2)|^2+2\left(\frac{A}{4(2\pi)^{\alpha+1}}\right)^2(n_1+1)^{-2\alpha}\\ &\geqslant \sum_{\bar m_1\bar m_2\leqslant n_1}(C(K+m_1,q_k+m_2)+C(-K+m_1,-q_k+m_2))^2\\ &\quad +2\left(\frac{A}{4(2\pi)^{\alpha+1}}\right)^2(n_1+1)^{-2\alpha}\\ &\geqslant C(K,1)^2+C(-K,-1)^2+2\left(\frac{A}{4(2\pi)^{\alpha+1}}\right)^2(n_1+1)^{-2\alpha}\\ &\geqslant 2\left(\frac{A}{4(2\pi)^{\alpha+1}}\right)^2(n^{-2\alpha}n_1^{2\alpha}+(n_1+1)^{-2\alpha})\\ &\geqslant 4\left(\frac{A}{4(2\pi)^{\alpha+1}}\right)^2 n^{-\alpha}\left(\frac{n_1}{n_1+1}\right)^{\alpha'}\\ &\geqslant \left(\frac{A}{(4\pi)^{\alpha+1}}\right)^2 n^{-\alpha}.\end{aligned}$$

定理 7.2 证完.

**附记** 由定理 7.2 可知, 当 $\alpha>1$ 时, 除因子 $p^\epsilon$ 之外, 定理 7.1 已不允许作本质的改进了.

## §8. 第二类 Fredholm 积分方程的渐近解法

为简单起见, 我们用大写拉丁字母表示 $s$- 维空间的矢量. 现在, 我们来研究第二类 Fredholm 方程.

$$\varphi(P)=\lambda\int_{G_s}K(P,Q)\varphi(Q)\varphi Q+f(P) \tag{8.1}$$

的渐近解法问题, 此处 $f(P)\in H_s^\alpha(A)$ 及 $K(P,Q)\in H_{2s}^\alpha(A)$. 命

$$D(\lambda)=1+\sum_{\nu=1}^{\infty}\frac{(-1)^\nu}{\nu!}\lambda^\nu\int_{G_{\nu_s}}K\left[\begin{array}{c}P_1,\cdots,P_\nu\\P_1,\cdots,P_\nu\end{array}\right]dP_1,\cdots dP_\nu$$

表示 Fredholm 核, 此处

$$K\left[\begin{array}{c}P_1,\cdots,P_\nu\\Q_1,\cdots,Q_\nu\end{array}\right]=\det(K(P_i,Q_j))\quad(1\leqslant i,j\leqslant\nu),$$

又命

$$\Delta(\lambda)=\det\left(\delta_{ij}-\frac{\lambda}{n}K(M_i,M_j)\right)\quad(1\leqslant i,j\leqslant n).$$

我们假定 $D(\lambda)\neq 0$.

**定理 8.1** 若求积公式

$$\sup_{F\in H_s^\alpha(A)}\left|\int_{G_s}F(P)dP-\frac{1}{n}\sum_{j=1}^{n}F(M_j)\right|\leqslant Ac(\alpha,s)\varphi(n)\tag{8.2}$$

成立, 此处 $\varphi(n)=o(1)$, 又若 $\tilde\varphi(M_k)$ 表示线性方程组

$$\tilde\varphi(M_j)=\frac{\lambda}{n}\sum_{k=1}^{n}K(M_j,M_k)\tilde\varphi(M_k)+f(M_j)(1\leqslant j\leqslant n)\tag{8.3}$$

的解, 则方程 (8.1) 的解可以写为:

$$\varphi(P)=f(P)+\sum_{j=1}^{n}K(P,M_j)\tilde\varphi(M_j)+O(\varphi(n)),$$

此处与“$O$”有关的常数仅依赖于 $\lambda$、$K$ 与 $f$.

**引理 8.1**[9] 命 $a_{ij}$ 为实数, $0\leqslant k\leqslant n$ 及

$$\tilde A_n(k)=\det(a'_{ij})\quad(1\leqslant i,j\leqslant n),$$

此处 $a'_{ii}=1+a_{ii}(1\leqslant i\leqslant k)$, 否则 $a'_{ij}=a_{ij}$. 又命 $A_n(k)$ 表示 $\tilde A(k)$ 在条件

$$|a_{ij}|\leqslant\gamma/n$$

之下的绝对值的上确界, 此处 $\gamma$ 为常数. 则存在常数 $\gamma_1$ 与 $\gamma_2$, 使对所有的正整数 $n$ 皆有

$$A_n(n-1)\leqslant\gamma_1/n\text{及}A_n(n)\leqslant\gamma_2.$$

**引理 8.2** [9] 若求积分式 (8.2) 成立, 则存在常数 $n_0 = n_0(\lambda, A, \alpha, s)$, 使当 $n > n_0$ 时有

$$|\Delta(\lambda)| \geqslant \frac{1}{2}|D(\lambda)|.$$

**引理 8.3** 方程 (8.1) 的解属于 $H_s^\alpha(A')$, 此处 $A'$ 为仅依赖于 $\lambda, K$ 与 $f$ 的常数.

**证** 由

$$\varphi(P) - f(P) = \lambda \int_{G_s} K(P, Q)\varphi(Q)\mathrm{d}Q$$

得

$$\begin{aligned}|(\varphi(P) - f(P))^{(\alpha\vartheta_1, \cdots, \alpha\vartheta_s)}| &\leqslant \sup_{Q \in G_s} |K(P, Q)^{(\alpha\vartheta_1, \cdots, \alpha\vartheta_s)}||\lambda| \int_{G_s} |\varphi(Q)|\mathrm{d}Q \\ &\leqslant A'',\end{aligned}$$

此处 $\vartheta_1, \cdots, \vartheta_s = 0$ 或 1, 所以 $\varphi(P) - f(P) \in H_s^\alpha(A'')$. 从而

$$\varphi(P) = f(P) + (\varphi(P) - f(P)) \in H_s^\alpha(A').$$

引理 8.3 证完.

**定理 8.1 的证明** 固定 $P$, 由引理 8.3 可知 $K(P, Q)\varphi(Q) \in H_s^\alpha(A')$, 所以

$$\varphi(P) = \frac{\lambda}{n}\sum_{k=1}^{n} K(P, M_k)\varphi(M_k) + f(P) + O(\varphi(n)),$$

从而

$$\varphi(M_j) = \frac{\lambda}{n}\sum_{k=1}^{n} K(M_j, M_k)\varphi(M_k) + f(M_j) + O(\varphi(n))(1 \leqslant j \leqslant n),$$

与 (8.3) 式相减, 我们得线性方程组

$$z_j = \sum_{k=1}^{n} a_{jk}z_k + b_j(1 \leqslant j \leqslant n),$$

此处

$$z_j = \varphi(M_j) - \tilde{\varphi}(M_j), a_{jk} = \frac{\lambda}{n}K(M_j, M_k), b_j = O(\varphi(n)).$$

命 $\Delta_k(\lambda)$ 表示将 $\Delta(\lambda)$ 的第 $k$ 列

$$\left(-\frac{\lambda}{n}K(M_1, M_k), \cdots, 1 - \frac{\lambda}{n}K(M_k, M_k), \cdots, -\frac{\lambda}{n}K(M_n, M_k)\right)'$$

换为

$$(b_1, \cdots, b_n)'$$

所得的行列式, 则得

$$z_j = \Delta_j(\lambda)/\Delta(\lambda)(1 \leqslant j \leqslant n).$$

当 $n$ 充分大时, 由引理 8.2 得

$$|\Delta(\lambda)| > \frac{1}{2}|D(\lambda)| > 0.$$

又由于

$$\left|\frac{\lambda}{n}K(M_j, M_k)\right| \leqslant |\lambda|A/n,$$

所以由引理 8.1 得

$$|\Delta_j| \leqslant |b_j B_j| + \sum_{\substack{1\leqslant k\leqslant n\\ k\neq j}} |b_k B_k| \leqslant \gamma_2|b_j| + \frac{\gamma_1}{n}\sum_{k=1}^{n}|b_k| = O(\varphi(n)),$$

此处 $B_k$ 为 $b_k$ 在 $\Delta_j$ 中的余因子. 因此

$$z_j = O(\varphi(n))(1 \leqslant j \leqslant n).$$

定理 8.1 证完.

**命**

$$M_j = \left(\left\{\frac{j}{n}\right\}\left\{\frac{h_1 j}{n}\right\}, \cdots, \left\{\frac{h_q j}{n}\right\}\right) \quad (1 \leqslant j \leqslant n)$$

为由文献 [1] 的 (1.2) 式定义的点集贯, 则由文献 [1] 定理 4 及定理 8.1 得

**定理 8.2** 命 $s = r$, 则方程 (8.1) 的解可以写为

$$\varphi(P) = f(P) + \frac{\lambda}{n}\sum_{j=1}^{n} K(P, M_j)\tilde{\varphi}(M_j) + O(n^{-\frac{\alpha}{2}-\frac{\alpha}{2q}+\varepsilon}),$$

此处 $\tilde{\varphi}(M_j)$ 表示线性方程组 (8.3) 的解, 而与“$O$”有关的常数仅依赖于 $\lambda, K, f, \mathscr{R}_r$ 与 $\varepsilon$.

## §9. 定理 4 的证明

**定理 4 的证明** 方程

$$\varphi(x) = \int_0^x K(x, y)\varphi(y)\mathrm{d}y + f(x) \tag{9.1}$$

的解可以由 Neumann 级数

$$\varphi(x) = f(x) + \sum_{\nu=1}^{\infty}\varphi_\nu(x) \tag{9.2}$$

给出, 此处

$$\varphi_\nu(x)=\int_0^x\int_0^{x_1}\cdots\int_0^{x_{\nu-1}}G_\nu(x;x_1,\cdots,x_\nu)\mathrm{d}x_1\cdots\mathrm{d}x_\nu,$$
$$G_\nu(x;x_1,\cdots,x_\nu)=K(x,x_1)K(x_1,x_2)\cdots,K(x_{\nu-1},x_\nu)f(x_\nu).$$

因 $f\in H_1^\alpha(A)$ 与 $K(x,y)\in H_2^\alpha(A)$, 所以

$$G_\nu(x;x_1,\cdots,x_\nu)\in H_{\nu+1}^\alpha(2^{(\alpha+1)(\nu+1)}A^{\nu+1}),$$

因此

$$|\varphi_\nu(x)|\leqslant 2^{(\alpha+1)(\nu+1)}A^{\nu+1}\int_0^x\cdots\int_0^{x_{\nu-1}}\mathrm{d}x_1\cdots\mathrm{d}x_\nu\leqslant\frac{2^{(\alpha+1)(\nu+1)}A^{\nu+1}}{\nu!},$$

从而

$$\Big|\sum_{\nu=Q+1}^\infty\varphi_\nu(x)\Big|\leqslant\sum_{\nu=Q+1}^\infty\frac{2^{(\alpha+1)(\nu+1)}A^{\nu+1}}{\nu!}\leqslant\frac{c(A,\varepsilon)^Q}{Q!}$$
$$\leqslant c(A,\alpha,\varepsilon)p^{-\nu(\alpha)\alpha+\varepsilon}. \tag{9.3}$$

命

$$\tilde{G}_\nu(x,x_1,\cdots,x_\nu)=\frac{1}{p}\sum_{j=1}^pK\left(x,\frac{j}{p}\right)K\left(\frac{j}{p},\frac{aj}{p}\right)\cdots,K\left(\frac{a^{\nu-2}j}{p},\frac{a^{\nu-1}j}{p}\right)f\left(\frac{a^{\nu-1}j}{p}\right)$$
$$\times\sum_{\bar{m}_1\cdots\bar{m}_\nu\leqslant g}e^{-2\pi i(m_1+m_2a+\cdots+m_\nu a^{\nu-1})j/p}e^{2\pi i(m_1x_1+\cdots+m_\nu x_\nu)},$$

由定理 7.1 可知, 存在整数 $a$ 使

$$|\varphi_\nu(x)-\int_0^x\cdots\int_0^{x_{\nu-1}}\tilde{G}_\nu\mathrm{d}x_1\cdots\mathrm{d}x_\nu|$$
$$\leqslant\int_0^x\cdots\int_0^{x_{\nu-1}}|G_\nu-\tilde{G}_\nu|\mathrm{d}x_1\cdots\mathrm{d}x_\nu$$
$$\leqslant\left(\int_0^x\cdots\int_0^{x_{\nu-1}}dx_1\cdots dx_\nu\right)^{\frac{1}{2}}||G_\nu-\tilde{G}_\nu||_{L_2}$$
$$\leqslant A^{\nu+1}c(\alpha,\varepsilon)^\nu P^{-\nu(\alpha)\alpha+\varepsilon}\quad(1\leqslant\nu\leqslant Q). \tag{9.4}$$

故由 (9.2)~(9.4) 式即得定理 4.

**附记** 1. 本节所用的方法可以处理更为一般的方程

$$\varphi(x_1,\cdots,x_{s+l})=\int_0^1\cdots\int_0^1\int_0^{x_{s+1}}\cdots\int_0^{x_{s+l}}K(x_1,\cdots,x_{s+l},y_1,\cdots,y_{s+l})$$

$$\times \varphi(y_1, \cdots, y_{s+l}) \mathrm{d}y_1 \cdots \mathrm{d}y_{s+l} + f(x_1, \cdots, x_{s+l}).$$

此处 $s \geqslant 0, l \geqslant 1, f \in H^{\alpha}_{s+l}(A)$ 及 $K \in H^{\alpha}_{2s+2l}(A)$.

2. 当 $\lambda$ 充分小时, 本节所用的方法也可以处理第二类 Fredholm 型的积分方程.

3. 由文献 [1] 中 (1.2) 式定义的点集贯在此不能用, 此乃由于我们不知道 $c(\mathscr{R}_r, \varepsilon)$ 与 $r$ 的明确关系, 即显式, 此处 $c(\mathscr{R}r, \varepsilon)$ 为文献 [1] 定理 2 中的常数.

## 参 考 文 献


[1] 华罗庚, 王元, 论一致分布与近似分析 —— 数论方法 (I), 中国科学, 1973, 4, 339~357.

[2] коробов, Н. М., Приближенное вычисление кратных интегралов с помощь-ю методов теории чисел, *ДАН СССР*, 115 (1957), 6, 1062~1065.

[3] коробов, Н. М., О приближенном вычислении кратных интегралов, *ДАН*С ССР, 124 (1959), 6, 1207~1210.

[4] Weil, A,. On some exponential sums *PNAS, USA*, 34 (1948), 5, 204~207.

[5] Шахов, Ю. Н., О вычислении интегралов расгущей кратности, *жур*. выч. *Мат. и Мат*. *фи*3., 5 (1965), 5, 911~916.

[6] Шаховю, Ю. Н., О приближенном рещении многе-мерных линейных уравнений Волбтера II рода методом итераций, Поп. жур. выч. мат. И матфиз,. изл. Наука, Москва; 1964, 75~100.

[7] бахвалов, Н. С., Теоремы вложения для классов функций с несколькими ограниченными производными, *Вес.Мос.уН-та.*, 3(1963) 7~16.

[8] бахвалов, Н. С, Об оптималЬных оценках сходимости квадратурных процессов и метолов интегрирования типа Монте-Карло на классах функций, Доп к жур. выч. мат. и мат. физ; Изд. Наука. Москва, 1964, 5~63.

[9] Коробов, Н. М., Теоретико-числовые методы в приближенном анализе, физмат. Лит. Москва, 1963.

# 论一致分布与近似分析 —— 数论方法 (III)*

华罗庚　　王　元

(中国科学院数学研究所)

## 提　　要

本文中, 我们将研究由方程 $f(x)=0$ 的根的初等对称函数定义的一致分布点集贯, 此处 $f(x)$ 为一个 Pisot-Vijayaraghavan 数 (参看 Jacobi-Perron 算法) 的极小多项式, 我们求得这些点集的偏差估计, 然后将它们用于数值积分问题, 为了应用的目的, 本文给出了两个关于 $f(x)$ 的建议.

## §1. 导　　言

如果代数整数 $\omega>1$ 除其自身外, 所有共轭数都位于开单位圆之内, 则称 $\omega$ 为一个 Pisot-Vijayaraghavan 数 ($PV$ 数), 命 $\omega$ 为一个 $s(\geqslant 2)$ 次的 $PV$ 数, 它有共轭 $\omega(=\omega^{(1)}\omega_s^{(1)}),\omega^{(2)},\cdots,\omega^{(s)}$ 且满足

$$\omega>1,|\omega^{(2)}|\leqslant|\omega^{(3)}|\leqslant\cdots\leqslant|\omega^{(s)}|<1. \tag{1.1}$$

又假定 $\omega$ 适合既约方程

$$f(\omega)=0,f(x)=x^s-a_{s-1}x^{s-1}-\cdots-a_1x-a_0, \tag{1.2}$$

其中 $a_0,a_1,\cdots,a_{s-1}$ 为整数,

本文仍采用 [3,4] 中的规定与记号

**定理 1**　命 $\boldsymbol{b}=(b_0,b_1,\cdots,b_{s-1})\neq 0$ 为一个整矢量、命 $(Q_n)(=(Q_n^{(s)}))$ 为由下面递推公式

$$Q_i=b_i(0\leqslant i\leqslant s-1),$$

$$Q_n=a_{s-1}Q_{n-1}+a_{s-2}Q_{n-2}+\cdots+a_1Q_{n-s+1}+a_0Q_{n-s}(n\geqslant s) \tag{1.3}$$

定义的整数贯, 则对于任何给予的大数 $N$, 皆存在 $n_0=n_0(\omega,\boldsymbol{b},N)$ 使当 $n>n_0$ 时

$$|Q_n|>N \tag{1.4}$$

* 原载"Scientia Sinica", 18, 2, 1975, 184~198. 参看"科学通报" 19, 12, 1974, 559~560.

与

$$\left|\frac{Q_{n+1}}{Q_n}-\omega\right|<c(\omega,\boldsymbol{b})|Q_n|^{-1-\rho} \tag{1.5}$$

成立, 此处

$$\rho=\frac{\log|\omega^{(s)}|}{\log\omega}. \tag{1.6}$$

由定理 1 立即推出

$$\left|\frac{Q_{n+k}}{Q_n}-\omega^k\right|<c(\omega,\boldsymbol{b})|Q_n|^{-1-\rho}(1\leqslant k\leqslant s-1,n>n_0). \tag{1.7}$$

由于 $1,\omega,\cdots,\omega^{s-1}$ 在有理数域 $R$ 上是线性独立的, 所以由 [3] 中定理 6.1, 6.2(作一点修改) 与 8.1 及 [4] 中定理 4.1 得

**定理 2** 集合贯

$$\left(\left\{\frac{k}{Q_n}\right\}\left\{\frac{Q_{n+1}k}{Q_n}\right\},\cdots\left\{\frac{Q_{n+s-1}k}{Q_n}\right\}\right)(1\leqslant k\leqslant|Q_n|,n>n_0) \tag{1.8}$$

是一致分布的且有偏差

$$\varphi(|Q_n|)=c(\omega,\boldsymbol{b},\varepsilon)|Q_n|^{-\frac{1}{2}-\frac{\rho}{2}+\varepsilon}, \tag{1.9}$$

此处 $\varepsilon$ 为任意给予正数.

**定理 3** 假定 $\alpha>1$, 则

$$\begin{aligned}\sup_{f\in E_s^\alpha(C)}&\left|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_s)\mathrm{d}x_1,\cdots\mathrm{d}x_s\right.\\&\left.-\frac{1}{|Q_n|}\sum_{k=1}^{|Q_n|}f\left(\frac{k}{Q_n},\frac{Q_{n+1}k}{Q_n},\cdots,\frac{Q_{n+s-1}k}{Q_n}\right)\right|\\&\leqslant C\cdot c(\omega,\boldsymbol{b},\alpha,\varepsilon)|Q_n|^{-\frac{\alpha}{2}-\frac{\rho\alpha}{2}+\varepsilon},(n>n_0).\end{aligned} \tag{1.10}$$

**定理 4** 假定 $1\geqslant\alpha>0$, 则

$$\begin{aligned}\sup_{f\in E_s^\alpha(A)}&\left|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_s)\mathrm{d}x_1,\cdots\mathrm{d}x_s\right.\\&\left.-\frac{1}{|Q_n|}\sum_{k=1}^{|Q_n|}f\left(\frac{k}{Q_n},\frac{Q_{n+1}k}{Q_n},\cdots,\frac{Q_{n+s-1}k}{Q_n}\right)\right|\\&\leqslant A\cdot c(\omega,\boldsymbol{b},\alpha,\varepsilon)|Q_n|^{-\frac{\alpha}{2}-\frac{\rho\alpha}{2}+\varepsilon},(n>n_0).\end{aligned} \tag{1.11}$$

为了应用目的, 我们给出下面两个建议:

1) 取 $a_0 = a_1 = \cdots = a_{s-1} = 1$. 记方程

$$F(x) = x^s - x^{s-1} - \cdots - x - 1 = 0, \tag{1.12}$$

的最大实根为 $\eta(=\eta^{(1)} = \eta_s^{(1)})$ 及其他根为 $\eta^{(2)}, \cdots, \eta^{(s)}$, 命 $(F_n)(= (F_n^{(s)}))$ 表示由下面的递推公式

$$F_0 = F_1 = \cdots = F_{s-2} = 0, F_{s-1} = 1,$$

$$F_{n+s} = F_{n+s-1} + F_{n+s-2} + \cdots + F_{n+1} + F_n \quad (n \geqslant 0). \tag{1.13}$$

定义的整数贯、通常, $(F_n)$ 称为维数 $s$ 的广义 Fibonacci 贯, 由 Jacobi-Perron 算法理论可知

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \eta, \tag{1.14}$$

(见 [1] 第七章), 在本文中, 我们将证明 $\eta$ 是一个 $PV$ 数及

$$\left|\frac{F_{n+1}}{F_n} - \eta\right| < c(\eta) F_n^{-1-\frac{1}{2^s \log 2}-\frac{1}{2^{2s+1}}} \quad (n \geqslant s), \tag{1.5}$$

从而如果在定理 2, 3, 4, 中取 $(Q_n) = (F_n)$ 则 (1.9)、(1.10) 与 (1.11) 的右端分别变成 $c(\eta)F_n^{-\frac{1}{2}-\frac{1}{2^{s+1}\log 2}}$, $C\cdot c(\eta,\alpha)F_n^{-\frac{\alpha}{2}-\frac{\alpha}{2^{s+1}\log 2}}$ 与 $A\cdot c(\eta,\alpha)F_n^{-\frac{\alpha}{2}-\frac{\alpha}{2^{s+1}\log 2}}$.

2) 取 $a_0 = 1, a_1 = \cdots = a_{s-2} = 0$ 及 $a_{s-1} = L$, 此处 $L$ 为一个整数 $\geqslant 2$. 记方程

$$G(x) = x^s - Lx^{s-1} - 1 = 0 \tag{1.16}$$

的最大实根为 $\tau(= \tau^{(1)} = \tau_s^{(1)})$ 及其他根为 $\tau^{(2)}, \cdots, \tau^{(s)}$, 命 $G_n(= (G_n^{(s)}))$ 为由下面递推公式

$$\begin{aligned} &G_0 = G_1 = \cdots = G_{s-2} = 0, G_{s-1} = 1, \\ &G_{n+s} = LG_{n+s-1} + G_n \quad (n \geqslant 0) \end{aligned} \tag{1.17}$$

定义的整数贯, 在本文中, 我们将证明 $\tau$ 是一个 $PV$ 数及

$$\left|\frac{G_{n+1}}{G_n} - \tau\right| < c(\tau) G^{-1-\frac{1}{s-1}+\frac{2}{(s-1)L\log L}-\frac{1}{(s-1)L^{s+3}}} \quad (n \geqslant s) \tag{1.18}$$

从而, 如果在定理 2, 3, 4 中取 $(Q_n) = (G_n)$, 则 (1.9)、(1.10)、(1.11) 的右端分别变成 $c(\tau)G_n^{-\frac{1}{2}-\frac{1}{2(s-1)}+\frac{1}{(s-1)L\log L}}$, $C\cdot c(\tau,\alpha)G_n^{-\frac{\alpha}{2}-\frac{\alpha}{2(s-1)}+\frac{\alpha}{(s-1)L\log L}}$ 与 $A\cdot c(\tau,\alpha)$ $G_n^{-\frac{\alpha}{2}-\frac{\alpha}{2(s-1)}+\frac{\alpha}{(s-1)L\log L}}$.

本文中得到的结果比前文 [3,4] 用实分圆域 $\mathscr{R}_s$ 得到的对应结果略粗一点, 但对应于 [3,4] 中的 $(h_1, \cdots, h_s; n)$, 寻求 $(Q_n)$ 所需的初等运算量却减少了.

最后, 我们将给出用“Wang520”计算机算出的一些数值结果, 例如

$$\sup_{f\in E_4^2(c)}\left|\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_4)\mathrm{d}x_1\cdots,\mathrm{d}x_4 - \frac{1}{F_{19}}\sum_{k=1}^{F_{19}} f\left(\frac{k}{F_{19}},\frac{F_{20}k}{F_{19}},\frac{F_{21}k}{F_{19}},\frac{F_{22}k}{F_{19}}\right)\right| < 0.0054C, \tag{1.19}$$

此处 $F_{19}=10671, F_{20}=20569, F_{21}=39648$ 与 $F_{22}=76424$.

## §2. 定理 1 的证明

1) 初等对称函数.

命

$$S_l=\omega^{(1)l}+\omega^{(2)l}+\cdots+\omega^{(s)l}, l=1,2,\cdots \tag{2.1}$$

悉知 $S_l$ 可以由下面的 Newton 递推公式

$$\begin{cases} S_1=a_{s-1}, \\ S_2=a_{s-1}S_1+2a_{s-2}, \\ \cdots\cdots \\ S_s=a_{s-1}S_{s-1}+a_{s-2}S_{s-2}+\cdots+a_1S_1+sa_0 \end{cases} \tag{2.2}$$

与

$$S_n=a_{s-1}S_{n-1}+a_{s-2}S_{n-2}+\cdots+a_1S_{n-s+1}+a_0S_{n-s} \tag{2.3}$$

来计算, 此处 $n>s$.

显然

$$|S_n-\omega^n|\leqslant\sum_{i=2}^{s}|\omega^{(i)}|^n<s-1,$$

所以

$$|s_n|<\omega^n+s-1<s\omega^n.$$

因此得有理逼近

$$\begin{aligned} &\left|\frac{S_{n+1}}{S_n}-\omega\right|=|S_{n+1}-\omega S_n||S_n|^{-1} \\ =&\left|\sum_{i=2}^{s}\omega^{(i)n}(\omega^{(i)}-\omega)\right||S_n|^{-1}<(\omega+1)(s-1)|\omega^{(s)}|^n|S_n|^{-1} \\ =&(\omega+1)(s-1)\omega^{-pn}|S_n|^{-1}<(\omega+1)(s-1)s^{\rho}|S_n|^{-1-\rho}(n>n_0) \end{aligned} \tag{2.4}$$

2) 假定初始矢量为 $\boldsymbol{b}=(0,\cdots,0,1)$. 命 $l_1,\cdots,l_s$ 为一个非负整数集及

$$\Delta(l_1,\cdots,l_s)=\begin{vmatrix} \omega^{(1)l_1}, & \omega^{(2)l_1}, & \cdots, & \omega^{(2)l_1} \\ & \cdots\cdots & & \\ \omega^{(1)l_s}, & \omega^{(2)l_s}, & \cdots, & \omega^{(s)l_s} \end{vmatrix} \tag{2.5}$$

命

$$P_l(s-1)=\frac{\Delta(l,s-2,\cdots,1,0)}{\Delta(s-1,s-2,\cdots,1,0)}, \tag{2.6}$$

则由于 $\Delta(s-1,s-2,\cdots,1,0)$ 可以整除 $\Delta(l,s-2,\cdots,1,0)$, 所以 $P_l(s-1)$ 为一个代数整数. 因 $P_l(s-1)$ 是 $\omega^{(1)},\cdots,\omega^{(s)}$ 的一个对称函数, 所以 $P_l(s-1)$ 是一个有理整数.

显然

$$\begin{aligned} P_0(s-1)=&P_1(s-1)=\cdots=P_{s-2}(s-1)\\ =&0, P_{s-1}(s-1)=1. \end{aligned} \tag{2.7}$$

由于当 $n\geqslant s$ 时,

$$\begin{aligned} \omega^{(i)n}=&\omega^{(i)n-s}\omega^{(i)s}\\ =&\omega^{(i)n-s}(a_{s-1}\omega^{(i)s-1}+a_{s-2}\omega^{(i)s-2}+\cdots+a_1\omega^{(i)}+a_0)\\ =&a_{s-1}\omega^{(i)n-1}+a_{s-1}\omega^{(i)n-2}+\cdots+a_1\omega^{(i)n-s+1}+a_0\omega^{(i)n-s}(0\leqslant i\leqslant s-1), \end{aligned}$$

所以

$$\begin{aligned} P_n(s-1)=&a_{s-1}P_{n-1}(s-1)+a_{s-2}P_{n-2}(s-1)+\cdots\\ &+a_1P_{n-s+1}(s-1)+a_0P_{n-s}(s-1)\quad(n\geqslant s). \end{aligned} \tag{2.8}$$

由 (2.6) 可知

$$P_n(s-1)=\sum_{i=1}^{s}\frac{\omega^{(i)n}}{\prod\limits_{j\neq i}(\omega^{(i)}-\omega^{(j)})},(n\geqslant 0). \tag{2.9}$$

所以得有理逼近

$$\left|\frac{P_{n+1}(s-1)}{P_n(s-1)}-\omega\right|<c(\omega)|P_n(s-1)|^{-1-\rho}\quad(n>n_0). \tag{2.10}$$

3) 定理 1 的证明:

命

$$P_l(i)=\frac{\Delta(s-1,\cdots,i+1,l,i-1,\cdots,0)}{\Delta(s-1,\cdots,0)}\ (0\leqslant i\leqslant s-1). \tag{2.11}$$

则

$$P_l(i)=\delta_{i,l},(0\leqslant i,l\leqslant s-1), \tag{2.12}$$

此处 $\delta_{il}$ 表示 Kronecker 记号及当 $n\geqslant s$ 时,

$$\begin{aligned}P_n(i)=&a_{s-1}P_{n-1}(i)+\alpha_{s-2}P_{n-2}(i)+\cdots\\&+a_1P_{n-s+1}(i)+a_0P_{n-s}(i)\quad(0\leqslant i\leqslant s-1),\end{aligned} \tag{2.13}$$

命

$$Q_l=b_0P_l(0)+b_1P_l(1)+\cdots+b_{s-1}P_l(s-1)\quad(l=0,1,\cdots). \tag{2.14}$$

则 $Q_l(l=0,1,\cdots)$ 适合

$$Q_i=b_i(0\leqslant i\leqslant s-1), \tag{2.15}$$

及当 $n\geqslant s$ 时,

$$Q_n=a_{s-1}Q_{n-1}+a_{s-2}Q_{n-2}+\cdots+a_1Q_{n-s+1}+a_0Q_{n-s}, \tag{2.16}$$

命 $W^{(j)i}$ 表示 $W^{(J)i}$ 在 $\Delta(s-1,\cdots,0)$ 中的余子式, 则

$$P_l(i)=\frac{1}{\Delta(s-1,\cdots,0)}\sum_{j=1}^{s}W_i^{(j)}\omega^{(j)l}\quad(0\leqslant i\leqslant s-1). \tag{2.17}$$

所以由 (2.14) 得

$$Q_l=\frac{1}{\Delta(s-1,\cdots,0)}\sum_{j=1}^{s}\omega^{(j)l}\sum_{i=1}^{s-1}b_iW_i^{(j)}(l=0,1,\cdots). \tag{2.18}$$

由于

$$\begin{aligned}W_i=&W_i^{(1)}=(-1)^{s-i+1}\begin{vmatrix}\omega^{(2)s-1}, & \cdots, & \omega^{(s)s-1}\\ & \cdots\cdots & \\ \omega^{(2)i+1}, & \cdots, & \omega^{(s)i+1}\\ \omega^{(2)i-1}, & \cdots, & \omega^{(s)i-1}\\ & \cdots\cdots & \\ 1, & \cdots, & 1\end{vmatrix}\\=&(-1)^{s-i+1}\sigma_{s-1-i}(\omega^{(2)},\cdots,\omega^{(s)})W_{s-1}\quad(0\leqslant i\leqslant s-1),\end{aligned} \tag{2.19}$$

此处 $\sigma_l(\omega^{(2)\cdots,\omega^{(s)}})=\sum\limits_{2\leqslant i_1<\cdots<i_l\leqslant s}\omega^{(i_1)}\cdots\omega^{(i_l)}$ 表示 $\omega^{(2)},\cdots,\omega^{(s)}$ 的初等对称函数, 所以易证

$$\sigma_l(\omega^{(2)},\cdots,\omega^{(s)})=g_l(\omega)\quad(0\leqslant l\leqslant s-1), \tag{2.20}$$

此处 $g_l(\omega)$ 是一个有理系数的 $l$ 次多项式, 其首项系数为 $\pm1$, 由于 $1,\omega,\cdots,\omega^{s-1}$ 是 $R(\omega)$ 的基底, 所以 $W_0,W_1,\cdots,W_{s-1}$ 在有理数域 $R$ 上线性独立. 特别地,

$$b_0W_0+\cdots+b_{s-1}W_{s-1}\neq 0. \tag{2.21}$$

由 (2.18) 可知

$$Q_l=\frac{(b_0W_0+\cdots+b_{s-1}W_{s-1})\omega^l}{\Delta(s-1,\cdots,0}+O(|\omega^{(s)}|^l), \tag{2.22}$$

此处与"$O$"有关的常数仅依赖于 $\omega$ 与 $\boldsymbol{b}$, 因此由 (2.21) 与 (2.22) 即得定理.

**附记 1** 为了实用目的, 我们建议取初始值 $Q_0=Q_1=\cdots=Q_{s-2}=0$ 及 $Q_{s-1}=1$(见 [1]).

**附记 2** 对于 $s=2,a_0=a_1=1$ 及 $\boldsymbol{b}=(0,1)$, 由 (2.18) 得熟知的公式

$$F_l^{(2)}=\frac{1}{\sqrt5}\left(\left(\frac{1+\sqrt5}{2}\right)^l-\left(\frac{1-\sqrt5}{2}\right)^l\right)\quad (l=0,1,2\cdots). \tag{2.23}$$

因此 (2.18) 可以看作是 (2.23) 的推广.

## §3. $\eta$ 的估计

**引 3.1** 我们有

$$2-\frac{1}{2^{s-1}}<\eta<2-\frac{1}{2^s} \tag{3.1}$$

及

$$|\eta^{(i)}|\leqslant\eta-1\quad (2\leqslant i\leqslant s) \tag{3.2}$$

证明引 3.1 之前, 我们先证明

**引 3.2** 若多项式

$$g(x)=a_sx^s+a_{s-1}x^{s-1}+\cdots+a_1x+a_0$$

的系数满足 $a_s\geqslant a_{s-1}\geqslant\cdots\geqslant a_1\geqslant a_0>0$, 则 $g(x)=0$ 没有模大于 1 的根. (见 [7].)

**证** 由于当 $|x|>1$ 时,

$$\begin{aligned}|(1-x)g(x)|\geqslant&a_s|x|^{s+1}-((a_s-a_{s-1})|x|^s+(a_{s-1}-a_{s-2})|x|^{s-1}\\&+\cdots+(a_1-a_0)|x|+a_0)>a_s|x|^s(|x|-1)>0,\end{aligned}$$

所以引理成立.

**引 3.1 的证明** 1) 记

$$Q(x) = (x-1)F(x) = x^{s+1} - 2x^s + 1.$$

则

$$\begin{aligned} Q\left(2-\frac{1}{2^s}\right) &= \left(2-\frac{1}{2^s}\right)^{s+1} - 2\left(2-\frac{1}{2^s}\right)^s + 1 \\ &= 1 - \left(2-\frac{1}{2^s}\right)^s \frac{1}{2^s} = 1 - \left(1-\frac{1}{2^{s+1}}\right)^s > 0 \end{aligned}$$

及

$$\begin{aligned} Q\left(2-\frac{1}{2^{s-1}}\right) &= \left(2-\frac{1}{2^{s-1}}\right)^{s+1} - 2\left(2-\frac{1}{2^{s-1}}\right)^s + 1 \\ &= 1 - \left(2-\frac{1}{2^{s-1}}\right)^s \frac{1}{2^{s-1}} = 1 - 2\left(1-\frac{1}{2^s}\right)^s \\ &= 1 - \left(s^{\frac{1}{s}} - \frac{1}{2^{s-\frac{1}{s}}}\right)^s. \end{aligned}$$

命

$$g(s) = 2^s - 1 - 2^{s-\frac{1}{s}}.$$

则

$$\begin{aligned} g'(s) &= 2^s \log 2 - 2^{s-\frac{1}{s}}\left(1+\frac{1}{s^2}\right)\log 2 \\ &= 2^s\left(1 - 2^{-\frac{1}{s}}\left(1+\frac{1}{s^2}\right)\right)\log 2. \end{aligned}$$

由于

$$2^s \geqslant \left(1+\frac{1}{s^2}\right)^{s^2},$$

即当 $s \geqslant 2$ 时, $g'(2) > 0$, 所以当 $s \geqslant 2$ 时, $g(s)$ 为递增函数, 因此

$$\begin{aligned} &2^s - 1 - 2^{s-\frac{1}{s}} > 0, \\ &2^{\frac{1}{s}} - \frac{1}{2^{s-\frac{1}{s}}} > 1. \end{aligned}$$

即得 $Q\left(2-\frac{1}{2^{s-1}}\right) < 0$. (3.1) 得证.

2) 命

$$F(x) = (x-\eta)R(x)$$

及

$$R(x) = x^{s-1} + \beta_{s-2}x^{s-2} + \cdots + \beta_1 x + \beta_0.$$

则

$$\begin{cases} -\eta\beta_0 = -1, \\ \beta_0 - \eta\beta_1 = -1, \\ \cdots\cdots \\ \beta_{s-2} - \eta = -1. \end{cases} \tag{3.3}$$

所以

$$\beta_0 = \frac{1}{\eta}, \beta_1 = \frac{\eta+1}{\eta^2}, \cdots, \beta_{s-3} = \frac{\eta^{s-3} + \eta^{s-4} + \cdots + \eta + 1}{\eta^{s-2}}$$
$$\beta_{s-2} = \frac{\eta^{s-2} + \eta^{s-3} + \cdots + \eta + 1}{\eta^{s-1}} = \eta - 1$$

命

$$\gamma_j \begin{cases} \beta_j/\beta_{j+1}, & 当0 \leqslant j \leqslant s-3; \\ \beta_j, & 当j = s-2. \end{cases}$$

由于当 $x \geqslant 0$ 时, $\frac{x}{x+1}$ 为 $x$ 的递增函数及 $\gamma_j = \frac{\eta^{j+1} + \eta^j + \cdots + \eta}{\eta^{j+1} + \eta^j + \cdots + \eta + 1}(0 \leqslant j \leqslant s-2)$, 所以

$$\gamma_{s-2} > \gamma_{s-3} > \cdots > \gamma_0. \tag{3.4}$$

命 $x = \beta_{s-2}y$, 则

$$\begin{aligned} \tilde{R}(y) = R(\beta_{s-2}y) =& \beta_{s-2}^{s-1}y^{s-1} + \beta_{s-2}^{s-1}y^{s-2} + \beta_{s-3}\beta_{s-2}^{s-3}y^{s-3} + \cdots \\ &+ \beta_1\beta_{s-2}y + \beta_0. \end{aligned}$$

由 (3.4) 得

$$\beta_{s-2}^{s-1} > \beta_{s-3}\beta_{s-2}^{s-3} > \cdots > \beta_1\beta_{s-2} > \beta_0. \tag{3.5}$$

所以由引 3.2 可知 $\tilde{R}(y) = 0$ 的根的模皆 $\leqslant 1$, 即 $R(x) = 0$ 的根的模皆 $\leqslant \beta_{s-2} = \eta - 1$.

引理证完.

**引 3.3** 当 $s = 2$ 时有 $|\eta^{(2)}| = \eta^{-1}$ 及当 $s = 3$ 时有 $|\eta^{(2)}| = |\eta^{(3)}| = \eta^{-\frac{1}{2}}$.

**证** 显然当 $s = 2$ 时有 $|\eta^{(2)}| = \eta^{-1}$, 当 $s = 3$ 时, 由于

$$x^3 - x^2 - x - 1 = (x - \eta)\left(x^2 + (\eta - 1)x + \frac{1}{\eta}\right)$$

及

$$(\eta - 1)^2 - \frac{4}{\eta} = \frac{\eta^3 - 2\eta^2 + \eta - 4}{\eta} = \frac{-\eta^2 + 2\eta - 3}{\eta} < 0,$$

所以 $\eta^{(2)},\eta^{(3)}$ 为一对共轭复数, 因此 $|\eta^{(2)}|=|\eta^{(3)}|=\eta^{-\frac{1}{2}}$.

引理证完.

## §4. $\tau$ 的估计

**引 4.1** 我们有

$$L<\tau<L+L^{-(s-1)} \tag{4.1}$$

及

$$(L+L^{-\frac{1}{s-1}})^{-\frac{1}{s-1}}<|\tau^{(i)}|<(L-(L-1)^{-\frac{1}{s-1}})^{-\frac{1}{s-1}} \quad (2\leqslant i\leqslant s). \tag{4.2}$$

**证** 1) 由

$$G(L)=-1<0$$

及

$$G(L+L^{-(s-1)})=L^{-(s-1)}(L+L^{-(s-1)})^{s-1}-1>0.$$

立即得到 (4.1).

2) 命 $\chi$ 为方程

$$g(x)=x^s+Lx^{s-1}-1=0$$

在 (0,1) 中的实根, 由于在 (0,1) 中 $g'(x)\neq 0$, 所以 $\chi$ 为 $g(x)=0$ 在 (0,1) 中唯一的实根. 由

$$g(0)=-1<0$$

及

$$g(1)=L>0,$$

可知对于任何适合 $0<\delta<\chi$ 的 $\delta$ 皆有 $g(\chi-\delta)<0$. 所以在圆 $|x|=\chi-\delta$ 上有

$$1>|x^s+Lx^{s-1}|.$$

由 Rouché定理可知 1 及 $G(x)$ 在区域 $|x|<\chi-\delta$ 中的零点个数相同. 因此 $G(x)$ 在区域 $|x|<\chi-\delta$ 中没有零点. 命 $\delta\to 0$, 则可知 $G(x)=0$ 的根的模皆 $\geqslant\chi$. 由于

$$\begin{aligned} g((L+L^{-\frac{1}{s-1}})^{-\frac{1}{s-1}}) =&((L+L^{-\frac{1}{s-1}})^{-\frac{1}{s-1}}+L)(L+L^{-\frac{1}{s-1}})^{-1}-1\\ &<(L^{-\frac{1}{s-1}}+L)(L^{-\frac{1}{s-1}}+L)^{-1}-1=0, \end{aligned}$$

所以

$$|\tau^{(i)}|\geqslant\chi>(L+L^{-\frac{1}{s-1}})^{-\frac{1}{s-1}} \quad (2\leqslant i\leqslant s).$$

3) 假定 $L>2$, 命 $Q$ 为方程

$$h(x)=x^2-Lx^{s-1}+1=0$$

在 (0,1) 中的唯一实根, 由

$$h(0)=1>0$$

及

$$h(1)=-L+2<0,$$

可知对于任何适合 $0<\delta<1-Q$ 的 $\delta$ 皆有 $h(Q+\delta)<0$. 所以在圆 $|x|=Q+\delta$ 上有

$$|Lx^{s-1}|>|x^s+1|.$$

由 Rouché定理可知 $x^{s-1}$ 与 $G(x)$ 在区域 $|x|<Q+\delta$ 中的零点个数相同. 所以 $G(x)$ 在区域 $|x|<Q+\delta$ 中有 $s-1$ 个零点, 命 $\delta\to 0$, 则可知 $G(x)=0$ 在区域 $|x|\leqslant Q$ 中有 $s-1$ 个根, 由于

$$\begin{aligned}
&h((L-(L-1)^{-\frac{1}{s-1}})^{-\frac{1}{s-1}})\\
=&((L-(L-1)^{-\frac{1}{s-1}})^{\frac{1}{s-1}}-L)(L-(L-1)^{-\frac{1}{s-1}})^{-1}+1\\
=&(L-(L-1)^{-\frac{1}{s-1}})^{-1}((L-(L-1)^{-\frac{1}{s-1}})^{-\frac{1}{s-1}}-(L-1)^{-\frac{1}{s-1}})<0,
\end{aligned}$$

所以

$$|\tau^{(i)}|\leqslant Q<(L-(L-1)^{-\frac{1}{s-1}})^{-\frac{1}{s-1}}\quad(2\leqslant i\leqslant s)$$

4) 假定 $L=2$, 今往证方程

$$G(x)=x^s-2x^{s-1}-1=0$$

有 $s-1$ 个模 $<1$ 的根, 这等价往证方程 $H(y)=y^s+2y-1=0$ 有 $s-1$ 个模 $>1$ 的根. 命 $\delta>0$, 则由 Rouché定理可知 $H_\delta(y)=y^s+(2+\delta)y-1$ 与 $y$ 在区域 $|y|<1$ 中的零点个数相同, 所以 $H_\delta(y)$ 有 $s-1$ 个模 $\geqslant 1$ 的零点. 命 $\delta\to 0$, 则 $H(y)=0$ 有 $s-1$ 个模 $\geqslant 1$ 的根. 因 $H(y)=0$ 没有根满足 $|y|=1$, 所以 $H(y)=0$ 有 $s-1$ 个模 $>1$ 的根.

引理证完.

**引 4.2**　当 $s=2$ 时有 $|\tau^{(2)}|=\tau^{-1}$ 及当 $s=3$ 时有 $|\tau^{(2)}|=|\tau^{(3)}|=\tau^{-\frac{1}{2}}$.

**证**　当 $s=2$ 时显然有 $|\tau^{(2)}|=\tau^{-1}$. 当 $s=3$ 时, 由于

$$x^3-Lx^2-1=(x-\tau)\left(x^2+(\tau-L)x+\frac{1}{\tau}\right)$$

及

$$(\tau-L)^2-\frac{4}{\tau}=\frac{\tau^3-2L\tau^2+L^2\tau-4}{\tau}=\frac{-L\tau^2+L^2\tau-3}{\tau}<0,$$

所以 $\tau^{(2)},\tau^{(3)}$ 是一对共轭复数, 因此 $|\tau^{(2)}|=|\tau^{(3)}|=\tau^{-\frac{1}{2}}$.

引理证完.

## §5. 多项式的既约性

在本书中, 我们将证明两条关于多项式既约性定理.

命

$$\omega(x)=x^s+a_{s-1}x^{s-1}+\cdots+a_1x+a_0, \tag{5.1}$$

此处 $a_0,\cdots,a_{s-1}$ 为整数.

**定理 5.1** (谢庭藩与裴定一 [8]) 若

$$|a_1|>|a_0^{s-1}|+|a_{s-1}a_0^{s-2}|+\cdots+|a_2a_0|+1, a_0\neq 0, \tag{5.2}$$

则 $\omega(x)$ 在有理数域 $R$ 上是既约的.

**证** 由 (5.2) 可知

$$|a_1a_0|>|a_0^s|+|a_{s-1}a_0^{s-1}|+\cdots+|a_2a_0^2|+|a_0|.$$

所以由 Rouché定理可知 $\omega(x)$ 与 $x$ 在区域 $|x|<|a_0|$ 中有同样数目的零点. 因此 $\omega(x)$ 在区域 $|x|<|a_0|$ 中只有一个零点 $\vartheta$.

由 (5.2) 易见方程 $\omega(x)=0$ 没有模为 $|a_0|$ 的根. 若 $\omega(x)=u(x)v(x)$, 此处 $u(x)$ 与 $v(x)$ 为次数 $\geqslant 1$ 的整系数多项式及 $u(\vartheta)=0$, 则 $v(x)=0$ 的根的模皆 $>|a_0|$. 所以

$$|a_0|=|\omega(0)|=|u(0)v(0)|\geqslant|v(0)|>|a_0|,$$

此为矛盾.

定理证完.

**定理 5.2** 若 $|a_0|=1$ 及 $\omega(x)=0$ 只有一个模 $\geqslant 1$ 的根 $\vartheta$, 则 $w(x)$ 在有理数域 $R$ 上既约

**证** 若 $\omega(x)=u(x)v(x)$, 此处 $u(x)$ 与 $v(x)$ 为次数 $\geqslant 1$ 的整系数多项式, 及若 $u(\vartheta)=0$, 则 $v(x)=0$ 的根的模皆 $<1$, 所以 $|v(0)|<1$, 此为矛盾.

定理证完.

由引理 3.1 与 4.1 立即推出 $F(x)$ 与 $G(x)$ 在有理数域 $R$ 上是既约的, 于是得到

**系 5.1** $\eta$ 与 $\tau$ 都是 $PV$ 数.

**附记 1** 定理 1 分别改进了 Perron[5]) 的一个结果及 Bernstein(见 [1], 定理 12) 的一个结果.

**附记 2** 多项式

$$x^s - x^{s-1} - 1 (s = 2, 3, \cdots)$$

在有理数域 $R$ 上并非都是既约的, 例如

$$x^5 - x^4 - 1 = (x^2 - x + 1)(x^3 - x - 1).$$

(见 [1] 定理 11.)

## §6. $\eta$ 与 $\tau$ 的有理逼近

**引 6.1** 若 $n \geqslant s$, 则

$$\left|\frac{F_{n+1}}{F_n} - \eta\right| < c(\eta) F_n^{-1-\frac{1}{2^s\log 2}-\frac{1}{2^{2s+1}}}. \tag{6.1}$$

**证** 由引 3.1 可知

$$\rho = -\frac{\log|\eta|^{(s)}}{\log\eta} \geqslant -\frac{\log(\eta-1)}{\log\eta} \leqslant \frac{-\log\left(1-\frac{1}{2^s}\right)}{\log 2}$$
$$\geqslant \frac{1}{2^s\log 2} + \frac{1}{2^{2s+1}}.$$

所以由定理 1 及系 5.1 即得引理

**引 6.2** 若 $n \geqslant s$, 则

$$\left|\frac{G_{n+1}}{G_n} - \tau\right| < c(\tau) G_n^{-1-\frac{1}{s-1}+\frac{2}{(s-1)L\log L}-\frac{1}{(s-1)L^{s+3}}}. \tag{6.2}$$

**证** 由于

$$\begin{aligned}\log(L-(L-1)^{-\frac{1}{s-1}}) &= \log L + \log(1 - L^{-1}(L-1)^{-\frac{1}{s-1}})\\ &\geqslant \log L - L^{-1}(L-1)^{-\frac{1}{s-1}} - L^{-2}(L-1)^{-\frac{2}{s-1}}\\ &\geqslant \log L - L^{-1} - L^{-2}\end{aligned}$$

及

$$\log(L + L^{-(s-1)}) = \log L + \log(1 + L^{-s}) \leqslant \log L + L^{-s},$$

所以

$$\rho = -\frac{\log|\tau^{(s)}|}{\log\tau} \geqslant \frac{\log L - L^{-1} - L^{-2}}{(s-1)(\log L + L^{-s})}$$

$$\geqslant \frac{1}{s-1}\left(1-\frac{1}{L\log L}-\frac{1}{L^2\log L}\right)\left(1-\frac{1}{L^s\log L}\right)$$
$$\geqslant \frac{1}{s-1}\left(1-\frac{1}{L\log L}-\frac{1}{L^s\log L}-\frac{1}{L^2\log L}+\frac{1}{L^{s+1}\log^2 L}\right)$$
$$\geqslant \frac{1}{s-1}\left(1-\frac{2}{L\log L}+\frac{1}{L^{s+3}}\right).$$

由定理 1 及系 5.1 即得引理

**附记** 若我们在引 6.1 与 6.2 的证明中分别用引 3.3 与 4.2 代替引 3.1 与 4.1, 则 6.1 与 6.2 的右端, 当 $s=2$ 时可以改进为 $c(\eta)F_n^{(2)-2}$ 与 $c(\tau)G_n^{(2)-2}$ 而当 $s=3$ 时可以改进为 $c(\eta)F_n^{(3)-3/2}$ 与 $c(\tau)G_n^{(3)-3/2}$.

## §7. 例

命 $n$ 为整数 $\geqslant 1$ 及 $a_1,\cdots,a_s$ 为整数, 记

$$Q_n(f)=\int_0^1\cdots\int_0^1 f(x_1,\cdots,x_s)\mathrm{d}x_1\cdots\mathrm{d}x_s-\frac{1}{n}\sum_{k=1}^{n}f\left(\frac{a_1k}{n},\cdots,\frac{a_sk}{n}\right). \tag{7.1}$$

**定理 7.1** 我们有

$$\sup_{f\in E_s^2(C)}|Q_n(f)|\leqslant C(H(n;a_1,\cdots,a_s)-1), \tag{7.2}$$

此处

$$H(n;a_1,\cdots,a_s)=\begin{cases}\dfrac{1}{n}\left(\left(1+\dfrac{\pi^3}{3}\right)^s+2\displaystyle\sum_{k=1}^{\frac{n-1}{2}}\prod_{\nu=1}^{s}\left(1-\dfrac{\pi^2}{6}\right.\right.\\ \left.\left.+\dfrac{\pi^2}{2}\left(1-2\left\{\dfrac{a_\nu k}{n}\right\}\right)^2\right)\right) & \text{当} 2\nmid n;\\ \dfrac{1}{n}\left(\left(1+\dfrac{\pi^2}{3}\right)^s+\left(1-\dfrac{\pi^2}{6}\right)^\mu\left(1+\dfrac{\pi^2}{3}\right)^{s-\mu}\right.\\ \left.+2\displaystyle\sum_{k=1}^{\frac{n}{2}-1}\prod_{\nu=1}^{s}\left(1-\dfrac{\pi^2}{6}+\dfrac{\pi^2}{2}\left(1-2\left\{\dfrac{a_\nu k}{n}\right\}\right)^2\right)\right), & \text{当} 2|n,\end{cases} \tag{7.3}$$

其中 $\mu$ 表示 $a_\nu(1\leqslant\nu\leqslant s)$ 中的奇数个数 (见 [9]).

**引 7.1** 我们有

$$\sum_{m=-\infty}^{\infty}\frac{e^{2\pi imx}}{\bar{m}^2}=1-\frac{\pi^2}{6}+\frac{\pi^2}{2}(1-2\{x\})^2.$$

**证**　由于

$$\int_0^1 \left(1-\frac{\pi^2}{6}+\frac{\pi^2}{2}(1-2x)^2\right) e^{-2\pi imx}\mathrm{d}x=\bar{m}^{-2},$$

所以引理得证.

**定理 7.1 的证明**　由引 7.1 可知

$$\begin{aligned}\sup_{f\in E_s^2(C)}|Q_n(f)|\leqslant&\frac{C}{n}\sum_{k=1}^{n}{\sum}'\frac{e^{2\pi i(a_1m_1+\cdots+a_sm_s)k/n}}{(\bar{m}_1\cdots\bar{m}_s)^2}\\ =&C\left(\frac{1}{n}\sum_{k=1}^{n}\prod_{\nu=1}^{s}\left(\sum_{m_\nu=-\infty}^{\infty}\frac{e^{2\pi ia_\nu m_\nu k/n}}{\bar{m}_\nu^2}\right)-1\right)\\ =&C\left(\frac{1}{n}\sum_{k=1}^{n}\prod_{\nu=1}^{s}\left(1=\frac{\pi^2}{6}+\frac{\pi^2}{2}\left(1-2\left\{\frac{a_\nu k}{n}\right\}\right)^2\right)-1\right)\\ =&C(H(n;a_1,\cdots,a_s)-1).\end{aligned}$$

定理证完.

取 $n=F_m^{(s)}, a_1=1$ 及 $a_\nu=F_{m+v-1}^{(s)}(2\leqslant\nu\leqslant s)$. 我们由 Wang520 计算机算出下面两张表

**表 1　$s=3,4$**

| $n$ | $H(n;a_1,a_2,a_3)$ | $n$ | $H(n;a_1,a_2,a_3,a_4)$ |
|---|---|---|---|
| 149 | 1.17442 | 401 | 1.36254 |
| 927 | 1.01102 | 2872 | 1.02416 |
| 1705 | 1.00480 | 10671 | 1.00540 |

**表 2　$s=5,6$**

| $n$ | $H(n;a_1,\cdots,a_5)$ | $n$ | $H(n;a_1,\cdots,a_5)$ |
|---|---|---|---|
| 13624 | 1.07428 | 29970 | 1.14458 |

**附记**　除本文所述的关于 Fibonacci 贯的两种推广外, $G.N.$ Raney[6] 还给出了另一种推广. 命 $Q=Q_n=(a_{ij})$, 此处 $a_{ij}=1$, 其中 $i+j\leqslant n+1$, 否则 $a_{ij}=0$. 命 $\varphi_{n,0}=(1,0,\cdots,0)'$ 及 $\varphi_{n,d+1}=Q\varphi_{n,d}(d=0,1,2,\cdots)$. 列矢量贯 $\varphi_{n,d}$ 称为 Fibonacci 贯的推广.

命 $D_n(\lambda)=\det(Q-\lambda I)$ 为 $Q$ 的特征方程, 则当 $n=\dfrac{p-1}{2}$ 时, $D_n(\lambda)$ 在有理数域 $R$ 上是既约的, 此处 $p$ 为素数 $\geqslant 5$, 但对于其他 $n$ 并非总是既约的. 当 $D_n(\lambda)$ 既约时, 则由本文的方法可得 $n$ 个整数贯 $\varphi_{n,d}(r)(r=1,2,\cdots,n;d=0,1,2,\cdots)$ 适

合下面的递推公式

$$\sum_{j=0}^{n}\begin{bmatrix} n-j+\left[\frac{1}{2}j\right] \\ \left[\frac{1}{2}j\right] \end{bmatrix}(-1)^{\frac{(n-j)(n-j-1)}{2}+j}\varphi_{n,k+j}(r)=0,$$

此处初始值分别为矩阵 $(\varphi_{n,0},\cdots,\varphi_{n,n-1})$ 的行矢量, 则得

$$(\varphi_{n,d}(1),\cdots,\varphi_{n,d}(n))'=\varphi_{n,d}.$$

由于一般说来, $D_n(\lambda)$ 不是一个 Pisot-Vijayaraghavan 数的极小多项式, 似乎不能期望 $\varphi_{n,d+1}(r)/\varphi_{n,d}(r)$ 迅速收敛于 $\lambda_1$, 此处 $\lambda_1$ 是矩阵 $Q$ 的有最大绝对值的特征根.

更正: 在文 [3], §3.1 中, 我们可以将 $\omega's$ 写成

$$\omega_l=2\cos\frac{2\pi g^l}{p}(1\leqslant l\leqslant r),$$

此处 $g$ 为 mod$p$ 的一个原根. 变换 $\sigma^j(1\leqslant j\leqslant q)$ 应该定义为

$$(\sigma^j)\quad \omega_l\to\omega_{l+j}(1\leqslant l\leqslant r).$$

## 参 考 文 献


[1] Bernstein, L.: *The Jacobi-Perron Algorithm, Its Theory and Application, Lec. Not. in Math;* Springer; (1971), 207.

[2] Cassels, J. W. S.: *An Introduction to Diophantine Approximation,* Camb. Univ. Pre., (1957).

[3] 华罗庚与王元 (Hua, Loo-Keng & wang Yuan): On uniform distribution and numerical analysis (I) (Number-theoretie method), *Sci, Sin.,* 16, (1973), 483~505.

[4] 华罗庚与王元 (Hua Loo-Keng & wang Yuan): On uniform distribution and numerical analysis (II) (Number-theoretic method), *Sci. Sin.,* 17 (1974), 331~348.

[5] Perron, O.: Grundlagen fur eine Theorie des Jacobischen Ketten-bruchalgorithmus, *Math. Ann,* 64, (1906), 1~76.

[6] Raney, G. M.: Generalization of the Fibonacci sequence to $n$ dimensions, *Can. J. of Math.,* 18 (1966), 332~349.

[7] Uspensky, J. V.: *Theory of equations,* MH book Com. Inc; (1948).

[8] 谢庭藩与裴定一 (Xie Ting-fan & Pei Ding-yi): On irreducibility of polyno-mials, *Kexue Tongbao* (to appear).

[9] See"*Applications of Number Theory to Numerical Analysis*", edited by S. K. Zaremba, Acad. Pre., (1972).

Hua Loo Keng   Wang Yuan

# Applications of Number Theory to Numerical Analysis

Springer-Verlag
Berlin Heidelberg New York
Science Press, Beijing
1981

Hua Loo Keng   Wang Yuan
Institute of Mathematics, Academia Sinica
Beijing
The People's Republic of China

*Revised edition of the original Chinese edition published by Science press Beijing 1978 as the first volume in the Academia Sinica's Series in Pure and Applied Mathematics.*

Distribution rights throughout the world, excluding The people's Republic of China, granted to Springer-Verlag Berlin Heidelberg New York

AMS Subject Classification (1980): 10-XX, 12Axx, 65-XX

ISBN 3-540-10382-1 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-10382-1 Springer-Verlag New York Heidelberg Berlin

Library of Congress Cataloging in Publication Dala. Hua, Loo Keng, 1911-. Applications of number theory to numerical analysis. Bibliography: p. 1. Numerical analysis. 2 Numbers, Theory of I. Wang Yuan, joint author. II. Title. QA297.H83.511. 80-22434


Typesetting: Science Press, Beijing, The People's Republic of China
Printed in Germany

Printing and binding: K. Triltsch, Würzburg

2141/3140-543210

# Preface

Owing to the developments and applications of computer science, mathematicians began to take a serious interest in the applications of number theory to numerical analysis about twenty years ago. The progress achieved has been both important practically as well as satisfactory from the theoretical view point. For example, from the seventeenth century till now, a great deal of effort was made in developing methods for approximating single integrals and there were only a few works on multiple quadrature until the 1950's. But in the past twenty years, a number of new methods have been devised of which the number theoretic method is an effective one.

The number theoretic method may be described as follows. We use number theory to construct a sequence of uniformly distributed sets in the $s$-dimensional unit cube $G_s$, where $s \geqslant 2$. Then we use the sequence to reduce a difficult analytic problem to an arithmetic problem which may be calculated by computer. For example, we may use the arithmetic mean of the values of integrand in a given uniformly distributed set of $G_s$ to approximate the definite integral over $G_s$ such that the principal order of the error term is shown to be of the best possible kind, if the integrand satisfies certain conditions. It is worth mentioning that the principal order of the error term of the Cartesian product formula for a classical single quadrature formula depends on and increases very rapidly with the dimension $s$. And though the error term in the Monte Carlo method is independent of $s$, it is in the sense of probability there, not in the usual sense of error. The number theoretic method may also be used to construct an approximate polynomial for the periodic function of $s$ variables and to treat the problems of the approximate solutions to integral equations and partial differential equations of certain types.

Many important methods and results in number theory, especially those concerning the estimation of trigonometrical sums and simultaneous Diophantine approximations as well as those of classical algebraic number theory, may be used to construct the uniformly distributed sequence in $G_s$. The fundamental concepts in the number theoretic method were advanced in 1957—1962. N. M. Korobov (1957) introduced

the $p$ set with the aid of the estimation of a complete exponential sum. Using the Sun Zi theorem (Chinese remainder theorem), J. H. Halton (1960) generalized the J. G. Van der Corput sequence. N. S. Bahvalov (1959) and C. B. Haselgrove (1962) introduced independently the *gp* (good point) set and N. M. Korobov (1959) and E. Hlawka (1962) proposed independently the *glp* (good lattice point) set. It was suggested by us in 1960 to define the uniformly distributed sets in $G_s$ by means of a set of independent units of the cyclotomic field by which an effective algorithm for obtaining a sequence of sets of rational numbers with the same denominators that approximate a basis of the field simultaneously was obtained, where the principal order of the error term is of the best possible kind. Perhaps, it is worth mentioning that the classical methods of best simultaneous rational approximations are ineffective from the view point of numerical analysis. In 1974, we proposed also a method for defining the uniformly distributed sequence by the recurrence formula defined by a PV (Pisot-Vijayaraghavan) number. In this book, we first illuminate these methods and give the estimates of the discrepancies of the sets so defined. Then we shall give various applications of them to numerical analysis and a table of *glp* sets as an appendix.

Aside from a knowledge of elementary number theory (see Hua Loo Keng [2]), we shall need several deeper theorems in number theory for which the references are given. Concerning the more extended methods and problems in the theory of uniform distribution and the theory of multiple quadrature, we refer the reader to monographs of S. Haber [1], Hsu Li Zhi and Zhou Yun Shi [1], L. Kuiper and H. Niederreiter [1], H. Niederreiter [5] and A. H. Stroud [1].

It is with great pleasure and gratitude that we acknowledge conversation and correspondence with professors Feng Kang, He Zuo Xiu, Hsu Li Zhi, Wang Guang Yin and Xu Zhong Ji and assistant professors Wan Qing Xuan, Wei Gong Yi and Xu Feng. We are indebted to professor B. J. Birch for his suggestion to refine the concept of effectiveness and to make a distinction between theoretical effectiveness and the effectiveness which can be attained by a computer. Finally, we are grateful to Science Press (Beijing) and Springer-Verlag for all their help and patience during the course of publication.

*January*, 1980

Hua Loo Keng
Wang Yuan

# Contents


**Preface**

**Chapter 1 Algebraic Number Fields and Rational Approximation** ··· 197

1.1. The units of algebraic number fields ··· 197

1.2. The simultaneous Diophantine approximation of an integral basis ··· 199

1.3. The real cyclotomic field ··· 201

1.4. The units of a cyclotomic field ··· 203

1.5. Continuation ··· 210

1.6. The Dirichlet field ··· 216

1.7. The cubic field ··· 219

Notes ··· 220

**Chapter 2 Recurrence Relations and Rational Approximation** ··· 221

2.1. The recurrence formula for the elementary symmetric function ··· 221

2.2. The generalization of $S_N$ ··· 222

2.3. PV number ··· 225

2.4. The roots of the equation $F(x) = 0$ ··· 227

2.5. The roots of the equation $G(x) = 0$ ··· 229

2.6. The roots of the equation $H(x) = 0$ ··· 232

2.7. The irreducibility of a polynomial ··· 234

2.8. The rational approximations of $\eta, \tau, \omega$ ··· 235

Notes ··· 238

**Chapter 3 Uniform Distribution** ··· 239

3.1. Uniform distribution ··· 239

3.2. Vinogradov's lemma ··· 242

3.3. The exponential sum and the discrepancy ··· 242

3.4. The number of solutions to the congruence ··· 245

3.5. The solutions of the congruence and the discrepancy ··· 247

3.6. The partial summation formula ··· 248

3.7. The comparison of discrepancies ··· 249

3.8. Rational approximation and the solutions of the congruence ··· 250

3.9. The rational approximation and the discrepancy ··· 251

3.10. The lower estimate of discrepancy ··· 254

Notes ··· 258

**Chapter 4 Estimation of Discrepancy** ········ 259
4.1. The set of equi-distribution ········ 259
4.2. The Halton theorem ········ 260
4.3. The $p$ set ········ 267
4.4. The $gp$ set ········ 270
4.5. The construction of good points ········ 272
4.6. The $\mathfrak{R}_s$ set ········ 273
4.7. The $\eta$ set ········ 275
4.8. The case $s = 2$ ········ 277
4.9. The $glp$ set ········ 279
Notes ········ 284
**Chapter 5 Uniform Distribution and Numerical Integration** ········ 286
5.1. The function of bounded variation ········ 286
5.2. Uniform distribution and numerical integration ········ 289
5.3. The lower estimation for the error term of quadrature formula ········ 295
5.4. The quadrature formulas ········ 297
Notes ········ 298
**Chapter 6 Periodic Functions** ········ 299
6.1. The classes of functions ········ 299
6.2. Several lemmas ········ 301
6.3. The relations between $H_s^\alpha(C)$, $Q_s^\alpha(C)$ and $E_s^\alpha(C)$ ········ 304
6.4. Periodic functions ········ 307
6.5. Continuation ········ 309
Notes ········ 315
**Chapter 7 Numerical Integration of Periodic Functions** ········ 316
7.1. The set of equi-distribution and numerical integration ········ 316
7.2. The $p$ set and numerical integration ········ 317
7.3. The $gp$ set and numerical integration ········ 322
7.4. The lower estimation of the error term for the quadrature formula ········ 326
7.5. The solutions of congruences and numerical integration ········ 328
7.6. The $glp$ set and numerical integration ········ 331
7.7. The Sarygin theorem ········ 335
7.8. The mean error of the quadrature formula ········ 337
7.9. Continuation ········ 340
Notes ········ 342
**Chapter 8 Numerical Error for Quadrature Formula** ········ 343
8.1. The numerical error ········ 343
8.2. The comparison of good points ········ 345

8.3. The computation of the $\eta$ set ··· 346
8.4. The computation of the $\mathfrak{R}_s$ set ··· 347
8.5. Examples of other $\mathcal{F}_s$ sets ··· 350
8.6. The computation of a *glp* set ··· 351
8.7. Several remarks ··· 356
8.8. Tables ··· 358
8.9. Some examples ··· 359
Notes ··· 362
**Chapter 9 Interpolation** ··· 363
9.1. Introduction ··· 363
9.2. The set of equi-distribution and interpolation ··· 364
9.3. Several lemmas ··· 368
9.4. The approximate formula of the function of $E_s^\alpha(C)$ ··· 370
9.5. The approximate formula of the function of $Q_s^\alpha(C)$ ··· 373
9.6. The Bernoulli polynomial and the approximate polynomial ··· 376
9.7. The $\Omega$ results ··· 380
Notes ··· 382
**Chapter 10 Approximate Solution of Integral Equations and Differential Equations** ··· 383
10.1. Several lemmas ··· 383
10.2. The approximate solution of the Fredholm integral equation of second type ··· 386
10.3. The approximate solution of the Volterra integral equation of second type ··· 391
10.4. The eigenvalue and eigenfunction of the Fredholm equation ··· 393
10.5. The Cauchy problem of the partial differential equation of the parabolic type ··· 396
10.6. The Dirichlet problem of the partial differential equation of the elliptic type ··· 398
10.7. Several remarks ··· 401
Notes ··· 402
Appendix Tables ··· 403
Bibliography ··· 413

# Chapter 1

# Algebraic Number Fields and Rational Approximation

## 1.1 The units of algebraic number fields

Let $Q$ denote the rational number field and $\alpha$ be an algebraic number of degree $s$. Then the algebraic number field $\mathscr{F}_s = Q(\alpha)$ is the field given by the polynomials in $\alpha$ of degree$< s$ with rational coefficients.

Let $s = r_1 + 2r_2$. Let $\alpha^{(1)}(= \alpha)$, $\alpha^{(2)}, \cdots, \alpha^{(s)}$ be the conjugates of $\alpha$, where $\alpha^{(1)}, \cdots, \alpha^{(r_1)}$ are real numbers, $\alpha^{(r_1+1)}, \cdots, \alpha^{(r_1+2r_2)}$ are complex numbers and $\alpha^{(r_1+r_2+1)} = \overline{\alpha^{(r_1+1)}}, \cdots, \alpha^{(r_1+2r_2)} = \overline{\alpha^{(r_1+r_2)}}$. For any $\xi \in \mathscr{F}_s$, we have

$$\xi^{(r_1+r_2+j)} = \overline{\xi^{(r_1+j)}}, \quad 1 \leqslant j \leqslant r_2.$$

In other words, there are at most $r_1 + r_2$ different absolute values

$$|\xi^{(1)}|, \cdots, |\xi^{(r_1)}|, |\xi^{(r_1+1)}|, \cdots, |\xi^{(r_1+r_2)}|$$

among the conjugates of $\xi$.

Suppose that $\omega_1 \cdots, \omega_s$ is an integral basis of $\mathscr{F}_s$. Form the matrix

$$\Omega = (\omega_j^{(i)}), \quad 1 \leqslant i, j \leqslant s,$$

the matrix

$$S = \Omega'\Omega = \left(\sum_{k=1}^{s} \omega_i^{(k)} \omega_j^{(k)}\right), \quad 1 \leqslant i, j \leqslant s$$

is called the fundamental matrix of $\mathscr{F}_s$. Clearly, it is a symmetric matrix with rational integer entries. The invariants of a fundamental matrix under the modular group are characteristic properties of the algebraic number field. The determinant det $S$ of $S$ is called the discriminant of the field.

Let $r = r_1 + r_2$. Let $\varepsilon_1, \cdots, \varepsilon_{r-1}$ be a set of units of $\mathscr{F}_s$. If

$$\det(\ln|\varepsilon_j^{(i)}|) \neq 0, \quad 2 \leqslant i \leqslant r, \quad 1 \leqslant j \leqslant r-1,$$

then the set of units $\varepsilon_1, \cdots, \varepsilon_{r-1}$ is called a set of independent units of $\mathscr{F}_s$. It follows by Dirichlet's unit theorem that there always exists a set of independent units in any algebraic number field. (CF. E. Landau [1].)

Hereafter, we use $c(f, \cdots, g)$ to denote a positive constant depending on $f, \cdots, g$ only, but not always with the same value. We use $\alpha$, $c$, $C, \cdots$ to denote absolute positive constants.

**Theorem 1.1** *Let $\gamma_1, \cdots, \gamma_r$ be a given set of real numbers satisfying*

$$\sum_{j=1}^{r_1} \gamma_j + 2 \sum_{j=r_1+1}^{r} \gamma_j = 0. \tag{1.1}$$

*Then there exists a unit $\eta \in \mathscr{F}_s$ such that*

$$c^{-1} e^{r_i} \leqslant |\eta^{(i)}| \leqslant c e^{r_i}, \quad 1 \leqslant i \leqslant r, \tag{1.2}$$

*where $c = c(\mathscr{F}_s)$.*

*proof.* Let $\varepsilon_1, \cdots, \varepsilon_{r-1}$ be a set of independent units of $\mathscr{F}_s$. Let

$$\xi^{(i)} = \varepsilon_1^{(i)a_1} \cdots \varepsilon_{r-1}^{(i)a_{r-1}}, \quad 1 \leqslant i \leqslant r.$$

Further let $c = e^a$, where

$$a = 2^{-1} \max_{1 \leqslant i \leqslant r} \left( \sum_{j=1}^{r-1} |\ln|\varepsilon_j^{(i)}|| \right). \tag{1.3}$$

Then

$$\prod_{i=1}^{r_1} |\xi^{(i)}| \prod_{j=1}^{r_2} |\xi^{(r_1+j)}|^2 = \prod_{k=1}^{r-1} \left( \prod_{i=1}^{r_1} |\varepsilon_k^{(i)}| \prod_{j=1}^{r_2} |\varepsilon_k^{(r_1+j)}|^2 \right)^{a_k} = 1. \tag{1.4}$$

Consider the system of linear equations

$$a_1 \ln|\varepsilon_1^{(i)}| + \cdots + a_{r-1} \ln|\varepsilon_{r-1}^{(i)}| = \gamma_i, \quad 1 \leqslant i \leqslant r. \tag{1.5}$$

Since

$$\det(\ln|\varepsilon_j^{(i)}|) \neq 0, \quad 2 \leqslant i \leqslant r, \quad 1 \leqslant j \leqslant r-1,$$

(1.5), except for the equation corresponding to $i = 1$, has a unique solution and it follows by (1.1) and (1.4) that this solution satisfies the equation for $i = 1$ in (1.5).

Let $b_k (1 \leqslant k \leqslant r-1)$ be the integers such that

$$|b_k - a_k| \leqslant \frac{1}{2}, \quad 1 \leqslant k \leqslant r-1.$$

Then we may define a unit $\eta$ of $\mathscr{F}_s$ by

$$\eta (= \eta^{(1)}) = \varepsilon_1^{b_1} \cdots \varepsilon_{r-1}^{b_{r-1}}.$$

From (1.3) and (1.5), we have

$$\begin{aligned}|\ln|\eta^{(i)}| - \ln|\xi^{(i)}|| &= |\ln|\eta^{(i)}| - \gamma_i| \\ &\leqslant \sum_{k=1}^{r-1} |b_k - a_k||\ln|\varepsilon_k^{(i)}|| \leqslant a.\end{aligned}$$

Hence we have (1.2). The theorem is proved.

For a real algebraic number field $\mathscr{F}_s$, set $\gamma_1 = \gamma$ and $\gamma_2 = \cdots = \gamma_r = \gamma'$. Then the condition (1.2) becomes

$$\gamma + (s-1)\gamma' = 0 \quad \text{or} \quad \gamma' = -\frac{\gamma}{s-1}.$$

Hence we have

**Theorem 1.2** *Let $\mathscr{F}_s$ be a real algebraic number field of degree $s$. Then for any given real number $\gamma$, there exists a unit $\eta \in \mathscr{F}_s$ such that*

$$c^{-1}e^{\gamma} \leqslant \eta \leqslant ce^{\gamma}$$

*and*

$$c^{-1}e^{-\frac{\gamma}{s-1}} \leqslant |\eta^{(i)}| \leqslant ce^{-\frac{\gamma}{s-1}}, \quad 2 \leqslant i \leqslant s,$$

*where $c = c(\mathscr{F}_s)$.*

Remark. We may assume that $\eta > 0$ in Theorem 1.2, otherwise we may use $-\eta$ instead of $\eta$.

## 1.2 The simultaneous Diophantine approximation of an integral basis

Let $\mathscr{F}_s$ be a real algebraic number field of degree $s$. Let $\omega_1, \cdots, \omega_s$ be an integral basis of $\mathscr{F}_s$. Take $\gamma = 1, 2, \cdots$ in Theorem 1.2. Then we may obtain a sequence of units $\eta_l (l = 1, 2 \cdots)$ such that

$$\eta_l > l, \quad |\eta_l^{(i)}| \leqslant c(\mathscr{F}_s)\eta l^{-\frac{1}{s-1}}, \quad 2 \leqslant i \leqslant s. \tag{1.6}$$

Put

$$n_l = \sum_{i=1}^{s} \eta_l^{(i)} \tag{1.7}$$

and

$$h_{l_j} = \sum_{i=1}^{s} \eta_l^{(i)} \omega_j^{(i)}. \tag{1.8}$$

Then $n_l$ and $h_{l_j} (1 \leqslant j \leqslant s)$ are rational integers and we have

**Theorem 1.3**

$$\left|\frac{h_{l_j}}{n_l} - \omega_j\right| \leqslant c(\mathscr{F}_s) n_l^{-1-\frac{1}{s-1}}, \quad 1 \leqslant j \leqslant s. \tag{1.9}$$

*Proof.* For simplicity, we omit the index $l$. By (1.6), (1.7) and (1.8), we have

$$n = \eta + O\left(\eta^{-\frac{1}{s-1}}\right) = \eta + O\left(n^{-\frac{1}{s-1}}\right) = \eta(1 + O\left(n^{-\frac{1}{s-1}}\right))$$

and

$$h_j = \eta\omega_j + O\left(\eta^{-\frac{1}{s-1}}\right) = \eta\omega_j(1 + O(n^{-1-\frac{1}{s-1}})).$$

Hence

$$\begin{aligned}\frac{h_j}{n} &= \omega_j(1 + O(n^{-1-\frac{1}{s-1}}))(1 + O(n^{-1-\frac{1}{s-1}}))^{-1} \\ &= \omega_j + O(n^{-1-\frac{1}{s-1}}), \quad 1 \leqslant j \leqslant s,\end{aligned}$$

where the constants implied by the symbol "$O$" depend on $\mathscr{F}_s$ only. The theorem is proved.

Now we shall give the expressions of $n$ and $h_j (1 \leqslant j \leqslant s)$. Let

$$\eta = \sum_{j=1}^{s} k_j \omega_j, \tag{1.10}$$

where $k_j (1 \leqslant j \leqslant s)$ are rational integers. Then we have

$$(\eta^{(1)}, \cdots, \eta^{(s)}) = (k_1, \cdots, k_s)\Omega'$$

and

$$(\eta^{(1)}, \cdots, \eta^{(s)})\Omega = (k_1, \cdots, k_s)S = (h_1, \cdots, h_s). \tag{1.11}$$

Let

$$\sum_{j=1}^{s} a_j \omega_j = 1, \tag{1.12}$$

where $a_j (1 \leqslant j \leqslant s)$ are rational integers. Then

$$n = \sum_{i=1}^{s} \eta^{(i)} \sum_{j=1}^{s} a_j \omega_j^{(i)} = \sum_{j=1}^{s} a_j \sum_{i=1}^{s} \eta^{(i)} \omega_j^{(i)} = \sum_{j=1}^{s} a_j h_j. \tag{1.13}$$

Hence we obtain the set of integers $(n, h_1, \cdots, h_s)$ corresponding to $\eta$ by (1.10), (1.11), (1.12) and (1.13).

*Remarks* 1. Theorem 1.3 is also true if the set $\omega_i (1 \leqslant i \leqslant s)$ is a basis of $\mathscr{F}_s$ and $\eta$ can be represented as a linear combination of $\omega_i' s$ with rational integer coefficients.

2. By Schmidt's theorem on simultaneous Diophantine approximation of algebraic numbers (Cf. W. M. Schmidt [2, 4]). We know that the estimate $c(\mathscr{F}_s)n_l^{-1-\frac{1}{s-1}}$ given in (1.9) is the best possible and it cannot be replaced even by $c(\mathscr{F}_s,\varepsilon)n_l^{-1-\frac{1}{s-1}-\varepsilon}$. Hereafter we use $\varepsilon$ to denote any pre-assigned positive number. But we have not yet considered here the best constant $c(\mathscr{F}_s)$ of (1.9). By the argument in the proof of Theorem 1.3, we know that it depends not only on the choice of $\eta_l$, but also on that of the integral basis. Let

$$|\overline{\omega}| = \max_{1\leqslant i,j\leqslant s}|\omega_j^{(i)}|.$$

Then the right hand side of (1.9) may be written as

$$|\overline{\omega}|c(\varepsilon_1,\cdots,\varepsilon_{r-1})n^{-1-\frac{1}{s-1}}.$$

3. Theorem 1.3 is not new. Our purpose here is to suggest a computational method for obtaining $(n,h_1,\cdots,h_s)$. For $s=2$, we can use continued fractions to treat the present problem, In the case of $s>2$, the situation is entirely different. The classical methods can only prove the existence of infinitely many sets of $(n,h_1,\cdots,h_s)$ satisfying (1.9). But they do not suggest any effective way (in the sense of numerical analysis) for finding $(n,h_1,\cdots,h_s)$. it is shown here that the problem of finding the sets of integers $(n,h_1,\cdots,h_s)$ is equivalent to the problem for finding a set of independent units in $\mathscr{F}_s$ and this requires only $c(\mathscr{F}_s)\ln n$ elementary operations for obtaining the set $(n,h_1,\cdots,h_s)$. Though Dirichlet's unit theorem is an existence theorem too, there are however many real algebraic number fields for which sets of independent units are known.

## 1.3 The real cyclotomic field

Let $m$ be an integer$\geqslant 5$ and $s=\dfrac{\varphi(m)}{2}$. The real cyclotomic field

$$\mathscr{F}_s = Q\left(\cos\frac{2\pi}{m}\right)$$

is an algebraic number field of degree $s$. The field has an integral basis

$$\omega_1=1,\quad \omega_l = 2\cos\frac{2\pi(l-1)}{m},\quad 2\leqslant l\leqslant s$$

(Cf. J. J. Liang [1]). Let $h_1(=1),h_2,\cdots,h_s$ be the integers satisfying $1\leqslant h<m/2$ and $(h,m)=1$. The transformation

$$\sigma_j:\omega_1\to\omega_1,\quad \omega_l\to 2\cos\frac{2\pi(l-1)h_j}{m},\quad 2\leqslant l\leqslant s$$

is an automorphism of the cyclotomic field $\mathscr{F}_s$ and the $s$ automorphisms

$$\sigma_1(=1), \sigma_2, \cdots \sigma_s$$

form the group of automorphism of the field. A number $\xi$ of $\mathscr{F}_s$ has $s$ conjugates under these automorphisms. Form the matrix

$$\Omega = (\sigma_i \omega_j), \quad 1 \leqslant i, j \leqslant s.$$

Then

$$S = \Omega' \Omega = (a_{ij}),$$

where $a_{11} = s$, $a_{1_j} = a_{j1} = C_m(j-1)(2 \leqslant j \leqslant s)$ and $a_{i_j} = C_m(i+j-2) + C_m(i-j)(2 \leqslant i, j \leqslant s)$ in which $C_m(k)$ denotes the trigonometric sum (Ramanujan sum)

$$C_m(k) = \sum_{(a,m)=1} e^{2\pi i ak/m},$$

which can be evaluated by the following lemma.

**Lemma 1.1** *$C_m(k)$ is a multiplicative function of $m$, i.e.,*

$$C_{m_1}(k) C_{m_2}(k) = C_{m_1 m_2}(k),$$

*if $(m_1, m_2) = 1$. And*

$$C_{p^l}(k) = \begin{cases} p^l - p^{l-1}, & \text{if} \quad p^l | k, \\ -p^{l-1}, & \text{if} \quad p^{l-1} \| k, \\ 0, & \text{it} \quad p^{l-1} \nmid k. \end{cases}$$

*Hereafter we always use $p$ to denote the prime number and $p^l \| b$ to denote $p^a | b$ but* $p^{a+1} \nmid b$ (Cf. Hua Loo keng [2], Chap. 7)

In particular, for the case $m = p$ , since

$$\sum_{l=1}^{s} 2 \cos \frac{2\pi l}{p} = -1,$$

we have an integral basis

$$\omega_l = 2 \cos \frac{2\pi}{p} g^l, \quad 1 \leqslant l \leqslant s,$$

of $\mathscr{F}_s = Q\left(\cos \dfrac{2\pi}{p}\right)$, where $g$ denotes a primitive root mod $p$. Hence

$$S = pI - M,$$

where $I$ is the $s \times s$ identity matrix and $M = (m_{i_j})$ is the matrix, where

$$m_{i_j} = 1 \quad (1 \leqslant i, j \leqslant s).$$

Suppose that $\varepsilon_1, \cdots, \varepsilon_{s-1}$ is the set of independent units of $Q\left(\cos\frac{2\pi}{m}\right)$ which will be given in next section. With the aid of $\varepsilon_i's$, we may construct by the method of §§1.1—1.2, a sequence of sets of integers $(n_l, h_{l1}, \cdots, h_{ls})(l = 1, 2, \cdots)$ such that

$$\left|\frac{h_{l_j}}{n_l} - 2\cos\frac{2\pi(j-1)}{m}\right| \leqslant c(\mathscr{F}_s)n_l^{-1-\frac{1}{s-1}}, \quad 2 \leqslant j \leqslant s, \tag{1.14}$$

where $h_{l1} = n_l$.

Hereafter we use the notations

$$c_{l1} = 1 \quad \text{and} \quad c_{lj} = h_{lj} \quad (2 \leqslant j \leqslant s).$$

*Remark* Since $Q\left(\cos\frac{2\pi}{5}\right) = Q(\sqrt{5})$, the sequence of sets $(n_l, c_{l1}, \cdots, c_{ls})(l = 1, 2, \cdots)$ which give the best simultaneous Diophantine approximation of the integral basis of $\mathscr{F}_s$ may be recognized as a generalization of the Fibonacci sequence and the integral basis of $Q\left(\cos\frac{2\pi}{m}\right)$ may be regarded as a generalization of the golden ratio in higher dimensional space.

## 1.4 The units of a cyclotomic field

We always use $p_1, p_2, \cdots$ to denote different prime numbers. Let $Z_m$ denote the multiplicative group of reduced residue classes modulo $m$. If the elements $h$ and $-h$ of $Z_m$ are identified, then we have the quotient group of $Z_m$ by $\{\pm 1\}$

$$Z_m/\{\pm 1\} = \{h_1, \cdots, h_s\}.$$

The $s$ characters of the group $Z_m/\{\pm 1\}$ are induced by the characters of the group $Z_m$ with $\chi(-1) = 1$ and they are denoted by

$$\chi_1, \cdots, \chi_s,$$

where $\chi_s$ denotes the principal character.

Set $h_{s+j} = h_j$. Let

$$\varepsilon(h) = \prod_{\substack{n|m, n>1 \\ p^l \| n \Rightarrow p^l \| m}} 2\sin\frac{\pi h}{n}, \quad (h, m) = 1$$

and

$$\eta_j = \varepsilon(h_{j+1})/\varepsilon(h_j), \quad 1 \leqslant j \leqslant s.$$

where $A \Rightarrow B$ means that $A$ implies $B$. The $\eta_j (1 \leqslant j \leqslant s)$ are all units of $\mathscr{R}_s$, since

$$\prod_{j=1}^{s} \eta_j = 1.$$

**Theorem 1.4** *$\eta_1, \cdots, \eta_{s-1}$ form a set of independent units of $\mathscr{R}_s$*

To prove the theorem, we shall need the following lemmas.

**Lemma 1.2**

$$\sum_{j=1}^{s} \chi(h_j) = \begin{cases} s, & \textit{if } \chi = \chi_s, \\ 0, & \textit{otherwise.} \end{cases}$$

*Proof.* Since

$$\sum_{(h,m)=1} \chi(h) = \begin{cases} 2s, & \text{if } \chi = \chi_s, \\ 0, & \text{otherwise,} \end{cases}$$

(Cf. Hua Loo Keng [2], Chap. 7), hence

$$\sum_{j=1}^{s} \chi(h_j) = \frac{1}{2} \sum_{(h,m)=1} \chi(h) = \begin{cases} s, & \text{if } \chi = \chi_s \\ 0, & \text{otherwise.} \end{cases}$$

The lemma is proved.

Let

$$\eta_j^{(i)} = \frac{\varepsilon(h_i h_{j+1})}{\varepsilon(h_i h_j)}, \quad 1 \leqslant i \leqslant s.$$

Then $\eta_j^{(1)} (= \eta_j), \eta_j^{(2)}, \cdots, \eta_j^{(s)}$ are all the conjugates of $\eta_j$. Let

$$C = (c_{ij}), \quad 1 \leqslant i, j \leqslant s-1,$$

where $c_{ij} = \ln|\eta_j^{(i)}| (1 \leqslant i, j \leqslant s-1)$.

**Lemma 1.3**

$$|\det C| = \prod_{j=1}^{s-1} \left| \sum_{i=1}^{s} \chi_j(h_i) \ln|\varepsilon(h_i)| \right|.$$

*Proof.* Clearly

$$|\det C| = |\det C^*|,$$

where $C^* = (c_{ij}^*), 1 \leqslant i, j \leqslant s$, in which

$$c_{ij}^* = \begin{cases} \ln|\varepsilon(h_i h_j)|, & \text{if } 1 \leqslant i \leqslant s-1, 1 \leqslant j \leqslant s, \\ 1, & \text{otherwise.} \end{cases}$$

Let

$$P = (\chi_j(h_i)), \quad 1 \leqslant i, j \leqslant s.$$

Then it follows by Lemma 1.2 that

$$|\det C \cdot \det P| = |\det C^* P| \\ = s\prod_{j=1}^{s-1}\left|\sum_{i=1}^{s}\chi_j(h_i)\ln|\varepsilon(h_i)|\right| |\det D|, \qquad (1.15)$$

where

$$D = (\overline{\chi}_j(h_i)), \quad 1 \leqslant i, j \leqslant s-1.$$

Since

$$s|\det D| = |\det D^*| = |\det \bar{P}|, \qquad (1.16)$$

where

$$D^* = \begin{pmatrix} D & \mathbf{I} \\ \mathbf{0} & s \end{pmatrix}$$

in which $\mathbf{0}$ and $\mathbf{I}$ denote the zero row vector and identity column vector respectively, and

$$|\det P| = |\det \bar{P}| = |\det \bar{P}' P|^{1/2} = s^{s/2}. \qquad (1.17)$$

The lemma follows by substituting (1.16) and (1.17) into (1.15).

Since

$$2\sum_{j=1}^{s}\chi(h_j)\ln|\varepsilon(h_j)| = \sum_{(a,m)=1}\chi(a)\ln|\varepsilon(a)|,$$

hence by the definition of the set of independent units, we have

**Lemma 1.4** *A necessary and sufficient condition that $\eta_1, \cdots, \eta_{s-1}$ form a set of independent units is that*

$$R_\chi = \sum_{(a,m)=1}\chi(a)\ln|\varepsilon(a)| \neq 0$$

*holds for any non-principal character $\chi$* mod $m$ *with* $\chi(-1) = 1$.

**Lemma 1.5**

$$\prod_{h=0}^{m-1} 2\sin\left(\frac{h\pi}{m}+\theta\right) = 2\sin m\theta.$$

*Proof*

$$\prod_{h=0}^{m-1} 2\sin\left(\frac{h\pi}{m}+\theta\right) = (-i)^m\prod_{h=0}^{m-1}(e^{(\frac{h\pi}{m}+\theta)i} - e^{-(\frac{h\pi}{m}+\theta)i}) \\ = (-i)^m e^{\frac{\pi i}{m}\sum\limits_{h=0}^{m-1}h - m\theta_i}\prod_{h=0}^{m-1}(e^{2\theta_i} - e^{-\frac{2\pi hi}{m}}) \\ = (-i)^m e^{\frac{\pi i(m-1)}{2} - m\theta_i}(e^{2m\theta_i} - 1)$$

$$= -i(e^{m\theta_i} - e^{-m\theta_i}) = 2\sin m\theta.$$

The lemma is proved.

**Lemma 1.6** *Suppose that* $0 < \theta < 1$. *Then*

$$\sum_{n=1}^{\infty} \frac{e^{2\pi in\theta}}{n} = -\ln(2\sin\pi\theta) + \left(\frac{\pi}{2} - \pi\theta\right) i.$$

*proof* Since

$$\sum_{n=1}^{\infty} \frac{r^n e^{2\pi in\theta}}{n} = -\ln(1 - re^{2\pi i\theta})$$

for $0 < r < 1$ and the series $\sum_{n=1}^{\infty} \frac{e^{2\pi in\theta}}{n}$ is convergent, hence, by Abel's theorem, we have

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{e^{2\pi in\theta}}{n} &= -\ln(1 - e^{2\pi i\theta}) \\ &= -\ln(e^{\pi i\theta}(e^{-\pi i\theta} - e^{\pi i\theta})) \\ &= -\ln((2\sin\pi\theta)e^{(\pi\theta - \frac{\pi}{2})i}) \\ &= -\ln(2\sin\pi\theta) + \left(\frac{\pi}{2} - \pi\theta\right) i.\end{aligned}$$

The lemma is proved.

**Lemma 1.7** *Let* $\chi$ *be a primitive character* mod $d$ *with* $\chi(-1) = 1$. *Then*

$$\sum_{(a,d)=1} \chi(a)\ln\left|2\sin\frac{\pi a}{d}\right| = -\tau(\chi)L(1,\bar{\chi}),$$

*where*

$$\tau(\chi) = \sum_{(r,d)=1} \chi(r)e^{2\pi ir/d}, \quad L(1,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n}.$$

*Proof.* Let

$$S(n,\chi) = \sum_{(r,d)=1} \chi(r)e^{2\pi inr/d}.$$

Then

$$\bar{\chi}(n)\tau(\chi) = S(n,\chi)$$

(Cf. Hua Loo Keng [2], Chap. 7) and so

$$\tau(\chi)L(1,\bar{\chi}) = \sum_{n=1}^{\infty} \frac{\bar{\chi}(n)\tau(\chi)}{n} = \sum_{n=1}^{\infty} \frac{1}{n} \sum_{(r,d)=1} \chi(r)e^{2\pi inr/d}$$

$$= \sum_{(r,d)=1} \chi(r) \sum_{n=1}^{\infty} \frac{e^{2\pi inr/d}}{n}$$

$$= \sum_{(r,d)=1} \chi(r) \left(-\ln\left|2\sin\frac{\pi r}{d}\right| + \left(\frac{\pi}{2} - \frac{\pi r}{d}\right) i\right)$$

$$= -\sum_{(r,d)=1} \chi(r)\ln\left|2\sin\frac{\pi r}{d}\right| - \frac{\pi i}{d}\sum_{(r,d)=1} \chi(r)r \tag{1.18}$$

by Lemma 1.6. Since $\chi(-1) = 1$, we have

$$\sum_{(r,d)=1} \chi(r)r = \sum_{(r,d)=1} \chi(d-r)(d-r) = -\sum_{(r,d)=1} \chi(r)r$$

and so

$$\sum_{(r,d)=1} \chi(r)r = 0.$$

Substituting into (1.18), the lemma follows.

**Lemma 1.18** *Let $m = p_1^{l_1} \cdots p_r^{l_r}$ and $d = p_1^{l_1'} \cdots p_r^{l_r'}$, where $l_i \geqslant 1$ and $0 \leqslant l_i' \leqslant l_i (1 \leqslant i \leqslant r)$. Let $\chi$ be a primitive character* mod $d$ *and also a character* mod $m$ *with $\chi(-1) = 1$. Then*

$$\sum_{(a,m)=1} \chi(a)\ln\left|2\sin\frac{\pi a}{m}\right| = -F(m_1)\tau(\chi)L(1,\bar{\chi}),$$

*where*

$$m_1 = \prod_{\substack{p|m \\ p\nmid d}} p, \quad F(m_1) = \prod_{p|m_1} (1-\chi(p)).$$

*Proof.* First, suppose that $m = dd'$ and $l_i' > 0 (1 \leqslant i \leqslant r)$. Then

$$r_\chi = \sum_{(a,m)=1} \chi(a)\ln\left|2\sin\frac{\pi a}{m}\right|$$

$$= \sum_{a_1=1}^{d'-1} \sum_{(a_2,d)=1} \chi(a_1 d + a_2)\ln\left|2\sin\frac{\pi(a_1 d + a_2)}{m}\right|$$

$$= \sum_{(a_2,d)=1} \chi(a_2) \sum_{a_1=0}^{d'-1} \ln\left|2\sin\left(\frac{\pi a_1}{d'} + \frac{\pi a_2}{m}\right)\right|$$

$$= \sum_{(a_2,d)=1} \chi(a_2)\ln\left|2\sin\frac{\pi a_2 d'}{m}\right|$$

$$= \sum_{(a,d)=1} \chi(a)\ln\left|2\sin\frac{\pi a}{d}\right| \tag{1.19}$$

by Lemma 1.5.

Next, suppose that $m = m_1m_2, m_1 = p_1^{l_1} \cdots p_j^{l_j}, m_2 = p_{j+1}^{l_{j+1}} \cdots p_r^{l_r}$ and $d|m_2$, where $1 \leqslant j < r$. Hence $\chi$ is also a character mod $m_2$ and

$$\begin{aligned} r_\chi &= \sum_{(a,m)=1} \chi(a) \ln \left| 2 \sin \frac{\pi a}{m} \right| \\ &= \sum_{(a_1,m_1)=1} \sum_{(a_2,m_2)=1} \chi(a_1m_2 + a_2m_1) \ln \left| 2 \sin \frac{\pi(a_1m_2 + a_2m_1)}{m} \right| \\ &= \sum_{(a_2,m_2)=1} \chi(a_2m_1) \sum_{a_1=1}^{m-1} \ln \left| 2 \sin \left( \frac{\pi a_1}{m_1} + \frac{\pi a_2}{m_2} \right) \right| \sum_{k|(a_1,m_1)} \mu(k) \\ &= \sum_{(a_2,m_2)=1} \chi(a_2m_1) \sum_{k|m_1} \mu(k) \sum_{l=1}^{\frac{m_1}{k}} \ln \left| 2 \sin \left( \frac{\pi lk}{m_1} + \frac{\pi a_2}{m_2} \right) \right|. \end{aligned}$$

It follows by Lemma 1.5 and (1.19) that

$$\begin{aligned} r_\chi &= \sum_{(a_2,m_2)=1} \chi(a_2m_1) \sum_{k|m_1} \mu(k) \ln \left| 2 \sin \frac{\pi a_2 m_1}{km_2} \right| \\ &= \sum_{k|m_1} \mu(k)\chi(k) \sum_{(a_2,m_2)=1} \overline{\chi}(k)\chi(a_2m_1) \ln \left| 2 \sin \frac{\pi a_2 m_1}{km_2} \right| \\ &= F(m_1) \sum_{(a,m_2)=1} \chi(a) \ln \left| 2 \sin \frac{\pi a}{m_2} \right| \\ &= F(m_1) \sum_{(a,d)=1} \chi(a) \ln \left| 2 \sin \frac{\pi a}{d} \right|. \end{aligned}$$

The lemma follows by Lemma 1.7.

**Lemma 1.9.** *Suppose that* $m = m_1m_2$, *where* $(m_1, m_2) = 1$. *Let* $\chi$ *be a character* mod $m$. *Then*

$$\sum_{(a,m)=1} \chi(a) \ln \left| 2 \sin \frac{\pi a}{m_2} \right| = \begin{cases} \varphi(m_1) \displaystyle\sum_{(a,m_2)=1} \chi(a) \ln \left| 2 \sin \frac{\pi a}{m_2} \right|, \\ \quad \text{if } \chi \text{ is a character mod } m_2, \\ 0, \quad \text{otherwise.} \end{cases}$$

*Proof.* Suppose that $\chi$ is not a character mod $m_2$. Then there exists an integer $t$ such that

$$(1 + tm_2, m) = 1, \chi(1 + tm_2) \neq 1.$$

Hence

$$\chi(1 + tm_2) \sum_{(a_1,m_1)=1} \chi(a_2m_1 + a_1m_2)$$

$$
\begin{aligned}
&= \sum_{(a_1,m_1)=1} \chi(a_2m_1 + a_1m_2(1+tm_2)) \\
&= \sum_{(a,m_1)=1} \chi(a_2m_1 + am_2)
\end{aligned}
$$

and so

$$\sum_{(a_1,m_1)=1} \chi(a_2m_1 + a_1m_2) = 0. \tag{1.20}$$

Suppose that $\chi$ is a character mod $m_2$. Then

$$\sum_{(a_1,m_1)=1} \chi(a_2m_1 + a_1m_2) = \varphi(m_1)\chi(a_2m_1). \tag{1.21}$$

Since

$$
\begin{aligned}
&\sum_{(a,m)=1} \chi(a)\ln\left|2\sin\frac{\pi a}{m_2}\right| \\
&= \sum_{(a_1,m_1)} \sum_{(a_2,m_2)=1} \chi(a_2m_1 + a_1m_2)\ln\left|2\sin\frac{\pi a_2 m_1}{m_2}\right| \\
&= \sum_{(a_2,m_2)=1} \ln\left|2\sin\frac{\pi a_2 m_1}{m_2}\right| \sum_{(a_1,m_1)=1} \chi(a_2m_1 + a_1m_2).
\end{aligned} \tag{1.22}
$$

The lemma follows by substituting (1.20) and (1.21) into (1.22).

The proof of Theorem 1.4 Let $\chi$ be any non-principal character mod $m$ with $\chi(-1) = 1$. Then $\chi$ is a primitive character mod $d$, where $d|m$ and $d \geqslant 3$. Hence it follows by Lemmas 1.8 and 1.9 that

$$
\begin{aligned}
R_\chi &= \sum_{(a,m)=1} \chi(a)\ln|\varepsilon(a)| \\
&= \sum_{\substack{n|m, n>1 \\ p^l \| n \Rightarrow p^l \| m}} \sum_{(a,m)=1} \chi(a)\ln\left|2\sin\frac{\pi a}{n}\right| \\
&= \sum_{\substack{n|m, n>1 \\ p^l \| n \Rightarrow p^l \| m}} \sum_{\chi(\operatorname{mod} n)} \varphi\left(\frac{m}{n}\right) \sum_{(a,n)=1} \chi(a)\ln\left|2\sin\frac{\pi a}{n}\right| \\
&= -\tau(\chi)L(1,\overline{\chi}) \sum_{\substack{n|m, n>1 \\ p^l \| n \Rightarrow p^l \| m \\ d|n}} \varphi\left(\frac{m}{n}\right) F(n_1)
\end{aligned} \tag{1.23}
$$

where $n_1 = \prod_{\substack{p^l \| n \\ p \nmid d}} p^l$.

Let $d' = \prod_{\substack{p^l \| n \\ p|d}} p^l$. Then $n = d'n_1$ and $(d', n_1) = 1$. Hence

$$\sum_{\substack{n|m, n>1 \\ p^l \| n \Rightarrow p^l \| m \\ d|n}} \varphi\left(\frac{m}{n}\right) F(n_1) = \varphi\left(\frac{m}{d'}\right) \sum_{\substack{n_1 | \frac{m}{d'} \\ p^l \| n_1 \Rightarrow p^l \| \frac{m}{d'}}} \frac{F(n_1)}{\varphi(n_1)}$$

$$= \varphi\left(\frac{m}{d'}\right) \prod_{p^l \| \frac{m}{d'}} \left(1 + \frac{1 - \chi(p)}{\varphi(p^l)}\right) \neq 0. \tag{1.24}$$

Since $\tau(\chi) \neq 0$ and $L(1, \overline{\chi}) \neq 0$ (Cf. Hua Loo Keng [2], Chap. 7 and Chap. 9), therefore $R_\chi \neq 0$ by (1.23). and (1.24). Hence the theorem follows by Lemma 1.4.

*Remark* In general, $\eta_1, \cdots, \eta_{s-1}$ is not a set of fundamental units (Cf. J. M. Masley and H. L. Montgomery [1] and W. Sinnott [1]).

## 1.5 Continuation

**Theorem 1.5** *Suppose that* $m = p_1^{l_1} \cdots p_r^{l_r}$, *where* $r \geqslant 2$ *and* $l_i \geqslant 1 (1 \leqslant i \leqslant r)$. *Then a necessary and sufficient condition that*

$$2 \sin \frac{\pi h}{m}, \quad (h, m) = 1, \quad 1 < h \leqslant \frac{m}{2} \tag{1.25}$$

*form a set of independent units is that the group* $Z_{m/p_i}^{l_i}$ *is generated by* $-1$ *and* $p_i$ *for any* $i$ *with* $1 \leqslant i \leqslant r$, *this is denoted by*

$$Z_{m/p_i^{l_i}} = \langle -1, p_i \rangle, \quad 1 \leqslant i \leqslant r. \tag{1.26}$$

To prove the theorem, we shall need the following lemmas.

**Lemma 1.10** *Suppose that* $n$ *is an integer* $> 1$. *Then*

$$\prod_{\substack{h=0 \\ (h,n)=1}}^{n-1} 2 \sin \frac{\pi h}{n} = \begin{cases} p, & \text{if } n = p^l, \\ 1, & \text{otherwise.} \end{cases}$$

*Proof.* By Lemma 1.5,

$$\ln \prod_{\substack{h=0 \\ (h,n)=1}}^{n-1} 2 \sin\left(\frac{h\pi}{n} + \theta\right) = \sum_{h=0}^{n-1} \ln\left(2 \sin\left(\frac{h\pi}{n} + \theta\right)\right) \sum_{d|(h,n)} \mu(d)$$

$$= \sum_{d|n} \mu(d) \sum_{k=0}^{\frac{n}{d}-1} \ln\left(2 \sin\left(\frac{kd\pi}{n} + \theta\right)\right)$$

$$= \sum_{d|n} \mu(d) \ln\left(2 \sin \frac{n\theta}{d}\right) = \ln \prod_{d|n} (\sin d\theta)^{\mu(\frac{n}{d})}.$$

Let $\theta \to 0$. Then the lemma follows.

We shall suppose hereafter that $m$ has at least two different prime factors. Let

$$H = (h_{ij}) = \left(\ln\left|2\sin\frac{\pi h_i h_j^*}{m}\right|\right), \quad 1 \leqslant i, j \leqslant s,$$

where $h_j^*$ denotes the inverse element of $h_j$ in $Z_m/\{\pm 1\}$. Then we have by Lemma 1.2 that

$$P'HP'^{-1} = (k_{ij}), \quad 1 \leqslant i, j \leqslant s,$$

where

$$\begin{aligned} k_{ij} &= \frac{1}{s}\sum_{r,t=1}^{s} \chi_i(h_r)\ln\left|2\sin\frac{\pi h_r h_t^*}{m}\right|\overline{\chi}_j(h_t) \\ &= \frac{1}{s}\sum_{r=1}^{s} \chi_i(h_r)\overline{\chi}_j(h_r)\sum_{t=1}^{s}\overline{\chi}_j(h_t h_r^*)\ln\left|2\sin\frac{\pi h_r h_t^*}{m}\right| \\ &= \delta_{i_j}\sum_{l=1}^{s}\chi_j(h_l)\ln\left|2\sin\frac{\pi h_l}{m}\right| \\ &= \frac{1}{2}\delta_{i_j}\sum_{(a,m)=1}\chi_j(a)\ln\left|2\sin\frac{\pi a}{m}\right|, \end{aligned}$$

in which $\delta_{i_j}$ denotes the Kronecker symbol, i.e., $\delta_{i_j} = 1$, if $i = j$ and $\delta_{i_j} = 0$ otherwise. This means that the $s$ eigenvalues of the matrix $H$ are

$$\lambda_i = \frac{1}{2}\sum_{(a,m)=1}\chi_i(a)\ln\left|2\sin\frac{\pi a}{m}\right|, \quad 1 \leqslant i \leqslant s.$$

Let $H_{i_j}$ denote the cofactor for $h_{i_j}$ in $H$. Then all the $H_{i_j}$'s are equal to each other, since the sums for every row and every column of $H$ are equal to zero. In particular, we have

$$H_{11} = H_{22} = \cdots = H_{ss}.$$

Denote the characteristic polynomial of $H$ by

$$\det(xI - H) = x^s + k_1 x^{s-1} + \cdots + k_{s-1}x + k_s.$$

Since

$$\lambda_s = \frac{1}{2}\sum_{(a,m)=1}\ln\left|2\sin\frac{\pi a}{m}\right| = 0$$

by Lemma 1.10, hence

$$k_{s-1} = (-1)^{s-1}(H_{11} + \cdots + H_{ss}) = (-1)^{s-1}sH_{11}$$

and

$$k_s = (-1)^s \lambda_1 \cdots \lambda_s = 0.$$

Since $H_{11} \neq 0$ means that (1.25) is a set of independent units, we have

**Lemma 1.11** *A necessary and sufficient condition that* (1.25) *is a set of independent units is that* 0 *is a simple root of the equation det* $(xI - H) = 0$, *i.e.,*

$$\lambda_\chi = \sum_{(a,m)=1} \chi(a) \ln \left| 2 \sin \frac{\pi a}{m} \right| \neq 0$$

*holds for any non-principal character* mod $m$ *with* $\chi(-1) = 1$.

The proof of Theorem 1.5. First, suppose that (1.26) holds. Let $\chi$ be a non-principal character mod $m$ and also a primitive character mod $d$ with $\chi(-1) = 1$. Then $d = p_1^{l_1'} \cdots p_r^{l_r'}$, where $0 \leqslant l_i' \leqslant l_i (1 \leqslant i \leqslant r)$. If $l_i' > 0 (1 \leqslant i \leqslant r)$, then

$$\lambda_\chi = -\tau(\chi) L(1, \overline{\chi}) \neq 0$$

by Lemma 1.8. If $l_1' = \cdots = l_j' = 0$ and $l_k' > 0 (j+1 \leqslant k \leqslant r)$, where $1 \leqslant j < r$, then by Lemma 1.8,

$$\lambda_\chi = -\prod_{i=1}^{j} (1 - \chi(p_i)) \tau(\chi) L(1, \overline{\chi}).$$

If there exists $p_i$ such that $\chi(p_i) = 1$, where $1 \leqslant i \leqslant j$, then it follows from $\chi(-1) = 1$ and (1.26) that $\chi(n) = 1$ for any $n \in Z_{m/p_i^{l_i}}$. Hence $\chi$ is a principal character mod $d$, since $d \left| \frac{m}{p_i^{l_i}} \right.$. This leads to a contradiction and so$\lambda_\chi \neq 0$. By Lemma 1.11, we know that (1.25) is a set of independent units.

Next, suppose that there exists an $i$ with $1 \leqslant i \leqslant s$ such that

$$Z_{m/p_i^{l_i}} \neq \langle -1, p_i \rangle.$$

Let $d = \frac{m}{p_i^{l_i}}$. Then there exists a primitive character mod $d^*$ such that $\chi(-1) = \chi(p_i) = 1$, where $d^* | d$. Hence $\lambda_\chi = 0$ by Lemma 1.8 and so (1.25) is not a set of independent units by Lemma 1.11. The theorem is proved.

From Theorem 1.5, we derive

**Theorem 1.6** *Suppose that* $m = p_1^{l_1} \cdots p_r^{l_r}$ *where* $r \geqslant 4$ *and* $l_i \geqslant 1 (1 \leqslant i \leqslant r)$ *or* $r \geqslant 3, p_1 = 2, l_1 \geqslant 3$ *and* $l_i \geqslant 1 (2 \leqslant i \leqslant r)$. *Then* (1.25) *is not a set of independent units.*

For example, (1.25) is independent for $m = 21$ and dependent for $m = 35$.

Now, we shall study the units of $\mathscr{R}_s$ for the case $m = p$. It follows by Theorem 1.4 that

$$\rho_j = \frac{\sin \frac{\pi}{p} g^{j+1}}{\sin \frac{\pi}{p} g^j}, \quad 1 \leqslant j \leqslant s-1$$

form a set of independent units of $\mathscr{R}_s$, where $g$ denotes a primitive root mod $p$. Clearly

$$\rho_j = \rho_{j+s}, \quad \rho_1 \cdots \rho_s = \pm 1.$$

We shall give the linear expression of $\rho_l$ by using the basis

$$\omega_l = 2\cos\frac{2\pi}{p}g^l, \quad l \leqslant l \leqslant s$$

of $\mathscr{R}_s$ as follows.

$$\begin{aligned}\rho_l &= ((e^{\frac{\pi i g^l}{p}})^g - (e^{-\frac{\pi i g^l}{p}})^g)(e^{\frac{\pi i g^l}{p}} - e^{-\frac{\pi i g^l}{p}})^{-1} \\ &= \sum_{m=0}^{g-1}(e^{\frac{\pi i g^l}{p}})^{g-1-m}(e^{-\frac{\pi i g^l}{p}})^m \\ &= \sum_{m=0}^{g-1} e^{\frac{\pi i g^l(g-1-2m)}{p}} = \sum_{m=0}^{g-1}\cos\frac{\pi g^l(g-1-2m)}{p}.\end{aligned}$$

First, suppose that $2|g$. Then

$$\rho_l = \sum_{m=0}^{\frac{1}{2}g-1} 2\cos\frac{\pi g^l(g-1-2m)}{p}.$$

Let

$$g-1-2m \equiv 2g^{e_m}(\mod p), \quad m = 0, 1, \cdots, \frac{1}{2}g - 1.$$

Then

$$\rho_l = \sum_{m=0}^{\frac{1}{2}g-1} \omega_{l+e_m}.$$

Since $g^s \equiv -1 \pmod p$, $l + e_m$ may be replaced by $l + e_m - ks$, where $ks < l + e_m \leqslant (k+1)s$. Hence

$$(\rho_1, \cdots, \rho_s) = (\omega_1, \cdots, \omega_s)M,$$

where $M$ is a circular matrix

$$M = \begin{pmatrix} a_1 & as \cdots a_2 \\ & \cdots \\ a_s & a_{s-1} \cdots a_1 \end{pmatrix},$$

is which $a_t = 1$, if $t$ equals $1 + e_m - s$ or $1 + e_m$ and $a_t = 0$ otherwise.

Next, suppose that $2 \nmid g$. Then

$$\rho_l = 1 + \sum_{m=0}^{\frac{1}{2}(g-3)} 2\cos\frac{\pi g^l(g-1-2m)}{p}.$$

Let

$$\frac{g-1}{2} - m \equiv g^{e_m} (\mathrm{mod} p), \quad m = 0, 1, \cdots, \frac{1}{2}(g-3).$$

Then

$$\rho_l = 1 + \sum_{m=0}^{\frac{1}{2}(g-3)} \omega_{l+e_m}.$$

Since

$$-1 = \sum_{l=1}^{s} \omega_l,$$

hence

$$(\rho_1, \cdots, \rho_s) = (\omega_1, \cdots, \omega_s)N,$$

where $N$ is also a circular matrix and its elements are 0 or $-1$.

By this way, any unit $\eta = \rho_1^{l_1} \cdots \rho_{s-1}^{l_{s-1}}$ can easily be expanded as a linear combination of $\omega_i$'s.

Let

$$x + x^{-1} = y \quad \text{or} \quad x = \frac{y \pm \sqrt{y^2 - 4}}{2}.$$

Then from the equation

$$x^{p-1} + x^{p-2} + \cdots + x + 1 = 0$$

we have the minimal equation of $\omega_l (1 \leqslant l \leqslant s)$

$$1 + \sum_{v=1}^{s} \left( \left( \frac{y + \sqrt{y^2 - 4}}{2} \right)^v + \left( \frac{y - \sqrt{y^2 - 4}}{2} \right)^v \right) = 0.$$

Hence

$$(-1)^s \prod_{l=1}^{s} \omega_l = 1 + \sum_{l=1}^{[\frac{s}{2}]} \frac{(-4)^l}{2^{2l-1}} = 1 + 2 \sum_{l=1}^{[\frac{s}{2}]} (-1)^l = (-1)^{[\frac{s}{2}]},$$

$$\prod_{l=1}^{s} \omega_l = (-1)^{[\frac{1}{2}(s+1)]} = (-1)^{\frac{p^2-1}{8}}$$

and so $\omega_l (1 \leqslant l \leqslant s)$ are all the units of $\mathscr{R}_s$.

**Theorem 1.7** *A necessary and sufficient condition that* $\omega_1, \cdots, \omega_{s-1}$ *form a set of independent units is that* 2 *is a primitive root* mod $p$ *or* 2 *belongs to the exponent* $s$ *and* $p \equiv 7$ (mod 8).

*Proof.* If 2 is a primitive root mod $p$, then by taking $g = 2$, we have

$$\rho_l = 2 \cos \frac{\pi 2^l}{p} = \omega_{l-1}, \quad 1 < l \leqslant s - 1$$

and

$$\rho_1 = 2\cos\frac{2\pi}{p} = 2\cos\frac{2^{s+1}\pi}{p} = \omega_s.$$

Now suppose that 2 is not a primitive root mod $p$. Let

$$2 \equiv g^l \pmod{p}.$$

Then $(l, p-1) > 1$ and

$$\omega_\mu = 2\cos\frac{2\pi}{p}g^\mu = \pm 2\cos\frac{\pi}{p}g^{\mu+l} = \frac{\sin\frac{\pi}{p}g^{\mu+2l}}{\sin\frac{\pi}{p}g^{\mu+l}}$$
$$= \rho_{l+\mu}\rho_{l+\mu+1}\cdots\rho_{2l+\mu-1}$$

and so

$$\omega_1\omega_{1+l}\cdots\omega_{1+(t-1)l} = (\rho_{l+1}\cdots\rho_{2l})(\rho_{2l+1}\cdots\rho_{3l})\cdots(\rho_{lt+1}\cdots\rho_{(l+1)t}).$$

The number of $\omega_i$'s in the product of the left hand side is $t$ and the number of $\rho_i$'s in the right hand side is $lt$.

Take $t = \frac{p-1}{(l,p-1)}$. If $(l, p-1) > 2$, then $t < s$ and $\frac{l(p-1)}{(l,p-1)}$ is a multiple of $p-1$ and so the right hand side is equal to $\pm 1$. Hence $\omega_1, \cdots, \omega_{s-1}$ is not a set of independent units. Suppose that $(l, p-1) = 2$. Then the exponent of 2 is $s$ and 2 is a quadratic residue mod $p$, i.e.,

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}} = 1.$$

Hence $p \equiv \pm 1 \pmod 8$ (Cf. Hua Loo Keng [2], Chap. 3).

First, suppose that $p \equiv 1 \pmod 8$. Then $l = 2m$, $2 \nmid m$ and $(m, p-1) = 1$, hence $g^m$ is a primitive root mod $p$ too. Without loss of generality, we may suppose that

$$2 \equiv g^2 \pmod{p}. \tag{1.27}$$

Hence there is an identity between $\frac{p-1}{4}$ of $\omega_i$'s

$$\omega_1\omega_3\cdots\omega_{2\frac{p-1}{4}-1} = \rho_3\rho_4\rho_5\rho_6\cdots\rho_1\rho_2 = \pm 1,$$

i.e., $\omega_1, \cdots, \omega_{s-1}$ is not a set of independent units.

Next, suppose that $p \equiv 7 \pmod 8$. Since $2\|(p-1)$, we may suppose that $2\|l$, otherwise $2\|(l+p-1)$. Hence we may assume that (1.27) holds also. If there exist $l_1, \cdots, l_s$ such that

$$\omega_1^{l_1}\omega_2^{l_2}\cdots\omega_s^{l_s} = \pm 1,$$

then

$$\pm 1 = (\rho_3\rho_4)^{l_1}(\rho_5\rho_6)^{l_2}\cdots(\rho_2\rho_3)^{l_s} = \rho_1^{l_{s-2}+l_{s-1}}\rho_2^{l_{s-1}+l_s}\cdots\rho_s^{l_{s-3}+l_{s-2}}.$$

Since the set $\rho_1, \cdots, \rho_{s-1}$ is independent and

$$\rho_1\cdots\rho_s = \pm 1,$$

we have

$$l_{s-2}+l_{s-1} = l_{s-1}+l_s = \cdots = l_{s-3}+l_{s-2}.$$

Since $2 \nmid s$, hence

$$l_1 = l_2 = \cdots = l_s.$$

if $l_s = 0$, then $l_1 = \cdots = l_{s-1} = 0$. Hence $\omega_1, \cdots, \omega_{s-1}$ form a set of independent units. The lemma is proved.

## 1.6 The Dirichlet field

Let $s = 2^t$. The field $\mathscr{D}_s = Q(\sqrt{p_1}, \cdots, \sqrt{p_t})$ is a real algebraic number field of degree $s$ which is called the Dirichlet field.

Consider the solution of Pell's equation

$$x^2 - p_{i_1}\cdots p_{i_k}y^2 = \pm 4, \quad x \geqslant 0, y \geqslant 0$$

with smallest $\dfrac{x}{2} + \dfrac{\sqrt{p_{i_t}\cdots p_{i_k}}\,y}{2}$, where $k \geqslant 1$ and $1 \leqslant i_1 < \cdots < i_k \leqslant t$ is any choice of $1, \cdots, t$. We order the $d_i = p_{i_1}\cdots p_{i_k}$ such that $d_i \equiv 1 \pmod 4$ for $1 \leqslant i \leqslant m$. Set

$$\varepsilon_i = \begin{cases} \dfrac{x_i}{2} + \dfrac{\sqrt{d_i}}{2}y_i, & x_i \equiv y_i \equiv 1(\text{mod}2), \quad \text{if} \quad 1 \leqslant i \leqslant m, \\ x_i + \sqrt{d_i}y_i, & \text{if} \quad m+1 \leqslant i \leqslant s-1. \end{cases}$$

Then $\varepsilon_1, \cdots, \varepsilon_{s-1}$ form a set of independent units of $\mathscr{D}_s$. We have also a basis of $\mathscr{D}_m$

$$\omega_1 = 1, \omega_2 = \varepsilon_1, \cdots, \quad \omega_{m+1} = \varepsilon_m, \omega_{m+2} = \sqrt{d_{m+1}}, \cdots, \quad \omega_s = \sqrt{d_{s-1}}.$$

The transformation

$$\sigma_{i_1\cdots i_k} : \sqrt{p_\nu} \to \begin{cases} -\sqrt{p_\nu}, & \text{if} \quad \nu = i_j (1 \leqslant j \leqslant k), \\ \sqrt{p_\nu}, & \text{otherwise} \end{cases}$$

is an automorphism of the field $\mathscr{D}_s$. $s-1$ transformations $\sigma_{i_1\cdots i_k}(k \geqslant 1, 1 \leqslant i_1 < \cdots < i_k \leqslant t)$ and the identity transformation $\sigma_0$ form the group of automorphism of the field. A number $\xi$ of $\mathscr{D}_s$ has its $s$ conjugates under these automorphisms.

**Lemma 1.12** *For any given integer $l \geqslant 1$, there exists a units $\eta_l$ of $\mathscr{D}_l$ such that*

$$\eta_l \geqslant \varepsilon_1^{(s-1)l} \tag{1.28}$$

*and*

$$|\eta_l^{(i)}| \leqslant c\eta_l^{-\frac{1}{s-1}}, \quad 2 \leqslant i \leqslant s, \tag{1.29}$$

*where $c = c(\mathscr{D}_s)$.*

*Proof.* Let $\Sigma_{i_1\cdots i_k}$ denote the number of $\varepsilon_\nu$'s which change into $\pm\varepsilon_\nu^{-1}$ under the transformation $\sigma_{i_1\cdots i_k}$. Suppose that $\varepsilon_\nu = X_\nu + \sqrt{d_\nu}Y_\nu$. Then $\sigma_{i_1\cdots i_k}\varepsilon_v = \begin{cases} \pm\varepsilon_\nu^{-1}, & \text{if } d_\nu \text{ is divisivle by odd number of } p_{i_j}(1 \leqslant j \leqslant k), \\ \pm\varepsilon_\nu, & \text{otherwise.} \end{cases}$ Since the number of $d_\nu$'s which possess $2r+1$ prime factors of $p_{i_j}(1 \leqslant j \leqslant k)$ and $h$ other prime factors is

$$\binom{k}{2r+1}\binom{t-k}{h},$$

therefore the number of $d_\nu$'s which have $2r+1$ prime factors of $p_{i_j}(1 \leqslant j \leqslant k)$ is equal to

$$\binom{k}{2r+1}\left(1+\binom{t-k}{1}+\cdots+\binom{t-k}{t-k}\right) = 2^{t-k}\binom{k}{2r+1}.$$

Hence

$$\Sigma_{i_1\cdots i_k} = \sum_{2r+1\leqslant k} 2^{t-k}\binom{t}{2r+1}.$$

Since

$$\sum_{i=0}^{k}(-1)^i\binom{k}{i} = (1-1)^k = 0$$

and

$$\sum_{i=0}^{k}\binom{k}{i} = (1+1)^k = 2^k,$$

hence

$$\sum_{2r\leqslant k}\binom{k}{2r} = \sum_{2r+1\leqslant k}\binom{k}{2r+1} = 2^{k-1}$$

and so

$$\Sigma_{i_1\cdots i_k} = 2^{t-1}. \tag{1.30}$$

Take $c = \varepsilon_1^{l_0 2^{t-1}}$, where $\varepsilon_1^{l_0} = \max_{2\leqslant j\leqslant s} \varepsilon_j$. For any given positive integer $l$, we may define the integers $l_j(2 \leqslant j \leqslant s)$ by

$$\varepsilon_1^l \leqslant \varepsilon_j^{l_j} < \varepsilon_1^{l+l_0}, \quad 2 \leqslant j \leqslant s-1. \tag{1.31}$$

Let

$$\eta_l = \varepsilon_1^l \varepsilon_2^{l_2} \cdots \varepsilon_{s-1}^{l_{s-1}}.$$

Then we have (1.28) by (1.31). (1.29) may be derived by (1.30) and (1.31) as follows.

$$\begin{aligned}|\eta_l^{(i)}| &< \varepsilon_1^{(l+l_0)(s-1-2^{t-1})-l2^{t-1}} = \varepsilon_1^{-l+l_0 2^{t-1}-l_0} \\ &= c\varepsilon_1^{-l-l_0} \leqslant c(\varepsilon_1^{-l}\varepsilon_2^{-l_2}\cdots\varepsilon_{s-1}^{-l_{s-1}})^{\frac{1}{s-1}} \\ &= c\eta_l^{-\frac{1}{s-1}} (2 \leqslant i \leqslant s).\end{aligned}$$

The lemma is proved.

Form the matrix

$$\Omega = (\omega_j^{(i)}), \quad 1 \leqslant i, j \leqslant s.$$

Then

$$S = \Omega' \Omega = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$

where $A = (a_{i_j})(1 \leqslant i, j \leqslant m+1)$ and $B = (b_{i_j})(1 \leqslant i, j \leqslant s-m-1)$ in which $a_{11} = 2^t$, $a_{ii} = 2^{t-2}(x_{i-1}^2 + d_{i-1}y_{i-1}^2)(2 \leqslant i \leqslant m+1)$, $a_{1j} = a_{j1} = 2^{t-1}x_{j-1}(2 \leqslant j \leqslant m+1)$, $a_{i_j} = 2^{t-2}x_{i-1}x_{j-1}(2 \leqslant i, j \leqslant m+1, i \neq j), b_{ii} = 2^t d_{m+i}(1 \leqslant i \leqslant s-m-1)$ and $b_{i_j} = 0(i \neq j)$.

Let $\eta_l(l = 1, 2, \cdots)$ be a sequence of units of $\mathscr{D}_s$ satisfying (1.28) and (1.29). Clearly $\eta_l$ may be expressed as a linear combination of $\omega_i$'s with rational integer coefficients

$$\eta_l = \sum_{i=1}^{s} k_{li}\omega_i.$$

Hence from

$$(h_{l1}, \cdots, h_{ls}) = (k_{l1}, \cdots, k_{ls})S,$$

we have

$$n_l = h_{l1} = 2^{t-1}\left(2k_{l1} + \sum_{i=1}^{m} x_i k_{l,i+1}\right),$$

$$h_{l_j} = \begin{cases} 2^{t-2}\left(2x_{j-1}k_{l1} + d_{j-1}y_{j-1}^2 k_{l_j} + \sum\limits_{i=2}^{m+1} x_{i-1}x_{j-1}k_{li}\right), & \text{if} \quad 2 \leqslant j \leqslant m+1, \\ 2^t k_{l_j} d_{j-1}, \quad \text{if} \quad m+2 \leqslant j \leqslant s \end{cases}$$

and the simultaneous Diophantine approximation of the basis of $\mathscr{D}_s$

$$\left|\frac{h_{lj}}{n_l} - \omega_j\right| < c(\mathscr{D}_s) n_l^{-1-\frac{1}{s-1}}, \quad 1 \leqslant i \leqslant s.$$

*Remark* The proof of Lemma 1.12 gives a simple algorithm for the computation of the sequence of units $\eta_l(l = 1, 2, \cdots)$ satisfying (1.28) and (1.29) which does not

involve the solution of the system of linear equation (1.5). But in practice the field $\mathscr{D}_s$ is not as convenient as $\mathscr{R}_s$, since the $\eta_l$ given by $\mathscr{D}_s$ increases too fast as $l$ increases and moreover $s$ is equal to $2^t$.

There are also many real algebraic number fields for which the sets of independent units are known (Cf. L. Bernstein, [1]).

## 1.7 The cubic field

Let $Q(\alpha)$ be a real cubic algebraic number field and let $\alpha$ satisfy the equation

$$x^3 - a_2x^2 - a_1x - a_0 = 0, \tag{1.32}$$

where $a_1$, $a_2$ are rational integers and $a_0 = \pm 1$. Suppose that (1.32) has only one real root and $\alpha > 1$. The field $Q(\alpha)$ has a basis

$$1, \alpha, \alpha^2$$

and a sequence of units $\eta_l = \alpha^l (l = 1, 2, \cdots)$ such that

$$|\eta_l^{(i)}| \leqslant \eta_l^{-\frac{1}{2}}, \quad i = 2, 3.$$

Hence we have

$$n_l = \alpha^l + \alpha^{(2)l} + \alpha^{(3)l}, \quad l = 1, 2, \cdots,$$
$$h_{lj} = \alpha^{l+j} + \alpha^{(2)l+j} + \alpha^{(3)l+j} = n_{l+j}, \quad j = 1, 2$$

and the simultaneous Diophantine approximation of the basis

$$\left| \frac{n_{l+j}}{n_l} - \alpha^j \right| \leqslant c(\alpha) n_l^{-\frac{3}{2}}, \quad j = 1, 2.$$

Here $n_l$ and $h_{l_j} (j = 1, 2)$ are taken from the same sequence of integers. It follows by (1.32) that

$$n_0 = 3, \quad n_1 = a_2, \quad n_2 = a_2^2 + 2a_1$$

and

$$\begin{aligned} n_l &= \sum_{i=1}^{3} \alpha^{(i)l} = \sum_{i=1}^{3} \alpha^{(i)l-3} \alpha^{(i)3} \\ &= \sum_{i=1}^{3} \alpha^{(i)l-3} (a_2 \alpha^{(i)2} + a_1 \alpha^{(i)} + a_0) \\ &= a_2 n_{l-1} + a_1 n_{l-2} + a_0 n_{l-3} \end{aligned}$$

for $l \geqslant 3$. Hence $n_l$ satisfies a simple recurrence relation.

For example, suppose that the minimal polynomial of $\alpha$ is

$$x^3 - x - 1 = 0.$$

Then $\alpha$ satisfies $1 < \alpha < 2$. Since

$$x^3 - x - 1 = (x - \alpha)(x^2 + \alpha x + \alpha^{-1})$$

and

$$\alpha^2 - 4\alpha^{-1} = (\alpha^3 - 4)\alpha^{-1} = (\alpha - 3)\alpha^{-1} < 0,$$

therefore $\alpha^{(2)}$ and $\alpha^{(3)}$ are conjugate complex numbers. Let

$$\eta_l = \alpha^l, \quad l = 1, 2, \cdots .$$

Then $n_l (l = 1, 2, \cdots)$ satisfy the recurrent formula

$$n_0 = 3, \quad n_1 = 0, \quad n_2 = 2, \quad n_l = n_{l-2} + n_{l-3} \quad (l \geqslant 3).$$

For real quadratic field $Q(\alpha)$, where $\alpha$ is a units$> 1$, the rational approximation of $\alpha$ thus obtained is in essence the continued fraction.

**Remark** It is easily seen that we may also obtain the less precise result of simultaneous Diophantine approximation of the $\omega_i$'s, if $\eta_l (l = 1, 2, \cdots)$ is a sequence of algebraic integers and

$$\sum_{i=2}^{s} |\eta_l^{(i)}| = o(\eta_l).$$

In particular, if we take $\eta_l = \alpha^l (l = 1, 2, \cdots)$, then the rational approximation of $\alpha$ so obtained is in essence the Jacobi-Perron algorithm. Here we may take $\alpha > 1$ and the absolute values of its conjugates are all less than 1 which then $\alpha$ is called a PV number. We shall discuss them in the next chapter.

**Notes**

Theorem1.4: Cf. K. Ramachandra, [1].

The other results: Cf. Hua Loo Keng and Wang Yuan [1, 4, 5, 6, 7, 8] and Hua Loo Keng, Wang Yuan and Pei Ding Yi [1].

# Chapter 2

# Recurrence Relations and Rational Approximation

## 2.1 The recurrence formula for the elementary symmetric function

Let $\mathscr{F}_s = Q(\alpha)$ be a real algebraic number field of degree $s$. We shall give in this chapter an algorithm for the simultaneous Diophantine approximation obtained by $\eta_l = \alpha^l (l = 1, 2, \cdots)$ which is essentially the Jacobi-Perron algorithm (Cf. L. Bernstein [1]). It yields less precise results but the computations of $n_l$ and $h_{l_j} (1 \leqslant j \leqslant s)$ are comparatively simple.

Let $\alpha$ satisfy the irreducible equation

$$f(x) = x^s - a_{s-1}x^{s-1} - \cdots - a_1 x - a_0 = 0, \tag{2.1}$$

where $a_{s-1}, \cdots, a_1, a_0$ are rational integers. Let

$$\alpha (= \alpha^{(1)}) > 1, \quad |\alpha^{(2)}| \leqslant \cdots \leqslant |\alpha^{(s)}| < 1.$$

An algebraic number with this property is called a PV number. Let

$$\rho = -\frac{\ln|\alpha^{(s)}|}{\ln\alpha}.$$

Then

$$|\alpha^{(i)}| \leqslant \alpha^{-\rho}, \quad 2 \leqslant i \leqslant s \tag{2.2}$$

and

$$0 < \rho \leqslant \frac{1}{s-1} - \frac{\ln|a_0|}{(s-1)\ln\alpha}, \tag{2.3}$$

since

$$|\alpha^{(s)}| = \frac{|a_0|}{\alpha|\alpha^{(2)} \cdots \alpha^{(s-1)}|} \geqslant \frac{|a_0|}{\alpha|\alpha^{(s)}|^{s-2}}.$$

Let $S_l$ denote the elementary symmetric functions of the roots of $f(x)$

$$S_l = \alpha^l + \alpha^{(2)l} + \cdots + \alpha^{(s)l}, \quad l = 0, 1, \cdots.$$

Then $S_l$'s are all rational integers. Without loss of generality, we may assume that $S_l > 0$, since $\alpha$ is a PV number. By (2.2), we have

$$|S_l - \alpha^l| \leqslant (s-1)|\alpha^{(s)}|^l = (s-1)\alpha^{-\rho l} \leqslant (s-1)s^\rho S_l^{-\rho}, \tag{2.4}$$

since

$$S_l \leqslant \alpha^l + (s-1) < s\alpha^l.$$

**Theorem2.1**

$$\left|\frac{S_{n+k}}{S_n} - \alpha^k\right| \leqslant c(\alpha) S_n^{-1-\rho}, \quad 1 \leqslant k \leqslant s-1.$$

Proof. From (2.4), we have

$$\begin{aligned}\frac{S_{n+k}}{S_n} &= \frac{\alpha^{n+k} + O(S_{n+k}^{-\rho})}{\alpha^n + O(S_n^{-\rho})} \\ &= \alpha^k(1 + O(S_n^{-1-\rho}))(1 + O(S_n^{-1-\rho}))^{-1} \\ &= \alpha^k + O(S_n^{-1-\rho}).\end{aligned}$$

The theorem is proved.

It is well-known that $S_l$ can be evaluated by Newton's formula

$$S_0 = s, \quad S_1 = a_{s-1}, \cdots, \quad S_{s-1} = a_{s-1}S_{s-2} + \cdots + a_1 S_0$$

and

$$S_n = a_{s-1}S_{n-1} + a_{s-2}S_{n-2} + \cdots + a_1 S_{n-s+1} + a_0 S_{n-s} \quad (n \geqslant s).$$

**Remark** It is easily seen from (2.3) that to take $\alpha$ to be a unit is more advantageous.

## 2.2 The generalization of $S_N$

Let $\xi$ be a number of $Q(\alpha)$ and

$$Q_n = \sum_{i=1}^{s} \xi^{(i)} \alpha^{(i)n}. \tag{2.5}$$

Then $Q_n$ is a rational number which is called the generalization of $S_n$. It also satisfies a recurrence relation. By (2.1), we have

$$\begin{aligned}Q_n &= \sum_{i=1}^{s} \xi^{(i)} \alpha^{(i)n-s} \alpha^{(i)s} \\ &= \sum_{i=1}^{s} \xi^{(i)} \alpha^{(i)n-s}(a_{s-1}\alpha^{(i)s-1} + \cdots + a_1\alpha^{(i)} + a_0)\end{aligned}$$

$$= a_{s-1}\sum_{i=1}^{s}\xi^{(i)}\alpha^{(i)n-1} + \cdots + a_1\sum_{i=1}^{s}\xi^{(i)}\alpha^{(i)n-s+1} + a_0\sum_{i=1}^{s}\xi^{(i)}\alpha^{(i)n-s}$$
$$= a_{s-1}Q_{n-1} + \cdots + a_1 Q_{n-s+1} + a_0 Q_{n-s} \tag{2.6}$$

for $n \geqslant s$. Hence the sequences $(Q_n)$ and $(S_n)$ differ only in their initial values.

Now we shall define $\xi$ such that $(Q_0, \cdots, Q_{s-1})$ is the given initial value. Choose

$$\Omega = (\alpha^{(i)j}), \quad 1 \leqslant i \leqslant s, \quad 0 \leqslant j \leqslant s-1.$$

Then

$$S = \Omega'\Omega$$

is a non-singular matrix with integer coefficients. From

$$(Q_0, \cdots, Q_{s-1}) = (\xi^{(1)}, \cdots, \xi^{(s)})\Omega,$$

we have

$$(Q_0, \cdots, Q_{s-1}) = (\xi^{(1)}, \cdots \xi^{(s)})\Omega^{-1}\Omega'\Omega$$

and so

$$(\xi^{(1)}, \cdots, \xi^{(s)}) = (Q_0, \cdots, Q_{s-1})S^{-1}\Omega'. \tag{2.7}$$

Hence $\xi \neq 0$, if $Q_0, \cdots, Q_{s-1}$ are not all equal to zero. And for any given integral initial vector $(Q_0, \cdots, Q_{s-1})$, we obtain a sequence of rational integers $Q_n (n = s, s+1, \cdots)$ by (2.6).

**Theorem 2.2** *Let $\boldsymbol{Q} = (Q_0, \cdots, Q_{s-1})$ be a non-zero integral vector. Then there exists a constant $c_1(\boldsymbol{Q}, \alpha)$ such that $|Q_n| > 1$ and that*

$$\left|\frac{Q_{n+k}}{Q_n} - \alpha^k\right| \leqslant c(\boldsymbol{Q}, \alpha)|Q_n|^{-1-\rho}, \quad 1 \leqslant k \leqslant s-1$$

*holds for $n > c_1(\boldsymbol{Q}, \alpha)$.*

*Proof* By (2.5) and (2.7), we have $|Q_n| > 1$ and

$$Q_n = \xi\alpha^n + O(|\alpha^{(s)}|^n) = \xi\alpha^n + O(\alpha^{-\rho n}) = \xi\alpha^n + O(|Q_n|^{-\rho})$$

for $n > c_1(\boldsymbol{Q}, \alpha)$, where the constant implied by the symbol "$O$" depends only on $\boldsymbol{Q}$ and $\alpha$. The theorem follows.

Let $\omega_1(=1), \omega_2, \cdots, \omega_s$ be any given basis of $Q(\alpha)$. Then

$$\omega_j = \sum_{k=1}^{s} t_{jk}\alpha^{k-1}, \quad 2 \leqslant j \leqslant s,$$

where $t_{jk}(2 \leqslant j \leqslant s, 1 \leqslant k \leqslant s)$ are rational numbers. Let

$$Q_n(j) = \sum_{k=1}^{s} t_{jk} Q_{n+k-1}, \quad 2 \leqslant j \leqslant s.$$

Clearly, $Q_n(j)$ satisfies also the recurrence relation

$$Q_n(j) = a_{s-1}Q_{n-1}(j) + \cdots + a_1 Q_{n-s+1}(j) + a_0 Q_{n-s}(j) \quad (n \geqslant s)$$

for $2 \leqslant j \leqslant s$, where the initial values $Q_0(j), \cdots, Q_{s-1}(j)$ are determined by $Q_0, \cdots, Q_{2s-2}$ and $t_{jk}(1 \leqslant k \leqslant s)$. From Theorem 2.2, we can derive

**Theorem 2.3** *Under the assumption of Theorem 2.2, the relation*

$$\left| \frac{Q_n(j)}{Q_n} - \omega_j \right| = O(|Q_n|^{-1-\rho}), \quad 2 \leqslant j \leqslant s$$

*holds for $n > c(\boldsymbol{Q}, \alpha)$, where the constant implied by the symbol "O" depends only on $Q$, $\alpha$ and $\omega_i$'s.*

We may also use the "Yang Hui triangular method" to prove Theorem 2.2. (Cf. Hua Loo Keng [3]). Suppose that $(Q_0, \cdots, Q_{s-1})$ is a given non-zero integral initial vector and that $Q_n (n \geqslant s)$ is a sequence of integers defined by (2.6). Then

$$\begin{aligned}
&(1 - a_{s-1}x - \cdots - a_1 x) \sum_{n=0}^{\infty} Q_n x^n \\
=&Q_0 + (Q_1 - a_{s-1}Q_0)x + \cdots \\
&+ (Q_{s-1} - a_{s-1}Q_{s-2} - \cdots - a_1 Q_0)x^{s-1} \\
=&P_{s-1}(x)(\text{say}).
\end{aligned}$$

Hence

$$\sum_{n=0}^{\infty} Q_n x^n = \frac{P_{s-1}(x)}{\prod_{i=1}^{s}(1 - \alpha^{(i)}x)} = \sum_{i=1}^{s} \frac{A_i}{1 - \alpha^{(i)}x}, \tag{2.8}$$

where

$$A_i = \frac{\alpha^{(i)s-1} P_{s-1}(\alpha^{(i)-1})}{\prod_{j \neq i}(\alpha^{(i)} - \alpha^{(j)})}, \quad 1 \leqslant i \leqslant s.$$

Expand the right hand of (2.8) as a power series and then we have

$$Q_n = \sum_{i=1}^{s} A_i \alpha^{(i)n} \tag{2.9}$$

by comparing the coefficients of $x^n$ of (2.8). Since $(Q_0, \cdots, Q_{s-1})$ is not a zero vector and $1, \alpha, \cdots, \alpha^{s-1}$ is a basis of $Q(\alpha)$, we have $A_i \neq 0 (1 \leqslant i \leqslant s)$ and so Theorem 2.2 follows by (2.9).

**Remarks** 1. In practical use, we often take the initial values $Q_0 = \cdots = Q_{s-2} = 0$, $Q_{s-1} = 1$ and the basis $1, \omega_2, \cdots, \omega_s$, where

$$\omega_j = \alpha^{j-1} - a_{s-1}\alpha^{j-2} - \cdots - a_{s-j+2}\alpha - a_{s-j+1}, \quad 2 \leqslant j \leqslant s.$$

(Cf. L. Bernstein [1]). Since

$$\omega_j = (a_{s-j}\alpha^{s-j} + \cdots + a_1\alpha + a_0)\alpha^{-s+j-1}, \quad 2 \leqslant j \leqslant s$$

by (2.1), we have

$$|\overline{\omega}| \leqslant s \max_{0 \leqslant j \leqslant s-1} |a_j|.$$

2. By (2.3), we have

$$0 < \rho \leqslant \frac{1}{s-1}.$$

Perron proved that the relation

$$\left|\frac{Q_{n+1}}{Q_n} - \alpha\right| < c(Q, \alpha)|Q_n|^{-1-\frac{1}{s-1}}$$

holds only for $s = 2$ (the continued fraction) and for $s = 3$ and $\alpha^{(2)}$, $\alpha^{(3)}$ are conjugate complex numbers (Cf. O. Perron [1]). Hence the results given here are rougher than the corresponding results of chapter 1. However compared with $(n_l, h_{l1}, \cdots h_{ls})$, the number of elementary operations required for obtaining $Q_n$ is decreased, since $Q_n$ satisfies a simple recurrence relation

## 2.3 PV numbers

Let $\mathscr{F}_s$ be a real algebraic number field of degree $s$. We shall show that it requires only $c(\mathscr{F}_s)$ elementary operations for obtaining a PV number $\alpha$ of degree $s$ in $\mathscr{F}_s$. Hence when $\mathscr{F}_s = Q(\alpha)$ we may obtain the simultaneous Diophantine approximation of the basis of $\mathscr{F}_s$ by the method stated in §2.2.

First, we mention two lemmas.

**Lemma2.1** *Let $s = r_1 + 2r_2$. Let $\xi_1, \cdots, \xi_s$ be $s$ linear forms of $x_1, \cdots, x_s$ with determinant $\Delta \neq 0$, where $\xi_1, \cdots, \xi_{r_1}$ are forms with real coefficients and the other forms have complex coefficients such that $\xi_{r_1+r_2+j} = \overline{\xi_{r_1+j}} (1 \leqslant j \leqslant r_2)$. Further let $\lambda_1, \cdots, \lambda_s$ be $s$ positive numbers satisfying $\lambda_{r_1+j} = \lambda_{r_1+r_2+j} (1 \leqslant j \leqslant r_2)$ and $\lambda_1 \cdots \lambda_s \geqslant \left(\frac{2}{\pi}\right)^{r_2} |\Delta|$. Then there exists a non-zero integer point such that*

$$|\xi_1| \leqslant \lambda_1, \cdots, |\xi_s| \leqslant \lambda_s$$

(Cf. Hua Loo Keng [2], Chap. 20).

**Lemma2.2** *Let $\alpha$ be a number of $\mathscr{F}_s$ satisfying the irreducible equation*

$$h(x) = 0, \quad \partial^0 h = l.$$

*Let*

$$g(x) = \prod_{i=1}^{s}(x - \alpha^{(i)}).$$

*Then $l/s$ and $g(x)$ is a polynomial with rational coefficients such that*

$$g(x) = ch(x)^{s/l},$$

*where $c$ is a rational number.* (Cf. Hua Loo Keng [2], Chap. 16).

Let $\omega_1, \cdots, \omega_s$ be a basis of $\mathscr{F}_s$, where the $\omega_i$'s are algebraic integers. Let $s = r_1 + 2r_2$ and

$$\alpha^{(i)} = \omega_1^{(i)} x_1 + \cdots + \omega_s^{(i)} x_s, \quad 1 \leqslant i \leqslant s, \tag{2.10}$$

where $\alpha^{(i)}(1 \leqslant i \leqslant r_1)$ have real coefficients and $\alpha^{(i)}(r_1 + 1 \leqslant i \leqslant s)$ have complex coefficients such that

$$\omega^{(r_1+r_2+j)} = \overline{\omega^{(r_1+j)}}, \quad 1 \leqslant j \leqslant r_2.$$

Let

$$\Omega = (\omega_j^{(i)}), \quad 1 \leqslant i, j \leqslant s.$$

Take

$$\lambda_1 = \left(\frac{2}{\pi}\right)^{r_2} |\Delta|(1-\varepsilon)^{-(s-1)}, \quad \lambda_2 = \cdots = \lambda_s = 1 - \varepsilon,$$

where $0 < \varepsilon < 1$ and $\Delta = \det\Omega$. Then it follows by Lemma 2.1 that there exists a non-zero integer point $(x_1, \cdots, x_s)$ such that

$$|\alpha| \leqslant \lambda_1 \quad |\alpha^{(i)}| \leqslant 1 - \varepsilon \left( \leqslant \left(\frac{2}{\pi}\right)^{\frac{r_2}{s-1}} |\Delta|^{\frac{1}{s-1}} \alpha^{-\frac{1}{s-1}} \right),$$

$$2 \leqslant i \leqslant s. \tag{2.11}$$

Hence $\alpha$ is a non-zero algebraic integer and we may assume that $\alpha > 0$. By

$$|N(\alpha)| = \alpha|\alpha^{(2)} \cdots \alpha^{(s)}| \geqslant 1,$$

we have

$$\alpha \geqslant (1-\varepsilon)^{-(s-1)} > 1 \tag{2.12}$$

and so $\alpha$ is a PV number of degree $s$ by (2.11), (2.12) and Lemma 2.2. From (2.10), we have

$$(x_1,\cdots,x_s)=(\alpha^{(1)},\cdots,\alpha^{(s)})\Omega'^{-1}$$

and so

$$\begin{aligned}|x_i| &\leqslant \left(\frac{2}{\pi}\right)^{r_2}\frac{W}{(1-\varepsilon)^{s-1}}+(s-1)(1-\varepsilon)\frac{W}{|\Delta|}\\ &= c(\mathscr{F}_s),\quad 1\leqslant i\leqslant s,\end{aligned} \tag{2.13}$$

where $W$ denotes the maximum of the absolute values of the cofactors of $\Delta$. Hence there is a non-zero integer point in the parallelpiped (2.13) satisfying (2.11) and so we obtain a PV number $\alpha$ of degree $s$ in $\mathscr{F}_s$.

## 2.4 The roots of the equation $F(x)=0$

Let $s\geqslant 2$. We denote the largest real root of the equation

$$F(x)=x^s-x^{s-1}-\cdots-x-1=0$$

by $\eta(=\eta^{(1)})$ and its other roots by $\eta^{(2)},\cdots,\eta^{(s)}$.

**Lemma2.3**

$$2-2^{-(s-1)}<\eta<2-2^{-s} \tag{2.14}$$

*and*

$$|\eta^{(i)}|\leqslant\eta-1,\quad 2\leqslant i\leqslant s. \tag{2.15}$$

To prove Lemma 2.3, we shall need

**Lemma2.4** *If the coefficients of the polynomial*

$$g(x)=a_sx^s+a_{s-1}x^{s-1}+\cdots+a_1x+a_0$$

*satisfy* $a_s\geqslant a_{s-1}\geqslant\cdots\geqslant a_1\geqslant a_0>0$, *then no root of the equation* $g(x)=0$ *has modulus greater than* 1.

*Proof* Since

$$\begin{aligned}|(1-x)g(x)| &\geqslant a_s|x|^{s+1}-((a_s-a_{s-1})|x|^s\\ &\quad+(a_{s-1}-a_{s-2})|x|^{s-1}+\cdots+(a_1-a_0)|x|+a_0)\\ &> a_s|x|^s(|x|-1)>0\end{aligned}$$

for $|x|>1$. The lemma follows

The proof of Lemma 2.3. Denote

$$Q(x)=(x-1)F(x)=x^{s+1}-2x^s+1.$$

Then

$$Q(2-2^{-s}) = (2-2^{-s})^{s+1} - 2(2-2^{-s})^s + 1$$
$$= 1 - 2^{-s}(2-2^{-s})^s = 1-(1-2^{-(s+1)})^s > 0$$

and

$$Q(2-2^{-(s-1)}) = (2-2^{-(s-1)})^{s+1} - 2(2-2^{-(s-1)}) + 1$$
$$= 1-2^{-(s-1)}(2-2^{-(s-1)})^s = 1-(2^{s^{-1}} - 2^{-s+s^{-1}})^s.$$

Let

$$R(x) = 2^s - 1 - 2^{s-s^{-1}}.$$

Then

$$R'(s) = 2^s\ln 2 - 2^{s-s^{-1}}(1-s^{-2})\ln 2$$
$$= 2^s(1-2^{-s^{-1}}(1-s^{-2}))\ln 2.$$

Since

$$2^s \geqslant (1+s^{-2})^{s^2},$$

i.e., $R'(s) > 0$ for $s \geqslant 2$, therefore $R(s)$ increases, if $s \geqslant 2$. Hence

$$2^s - 1 - 2^{s-s^{-1}} > 0,$$
$$2^{s^{-1}} - 2^{-s+s^{-1}} > 1$$

and so $Q(2-2^{-(s-1)}) < 0$. (2.14) is thus proved.

Let

$$F(x) = (x-\eta)f(x)$$

and

$$f(x) = x^{s-1} + \beta_{s-2}x^{s-2} + \cdots + \beta_1 x + \beta_0.$$

Then we have a system of linear equations

$$-\eta\beta_0 = -1,$$
$$\beta_0 - \eta\beta_1 = -1,$$
$$\cdots\cdots$$
$$\beta_{s-2} - \eta = -1.$$

Hence

$$\beta_0 = \frac{1}{\eta}, \quad \beta_1 = \frac{\eta+1}{\eta^2}, \cdots, \quad \beta_{s-3} = \frac{\eta^{s-3}+\eta^{s-4}+\cdots+\eta+1}{\eta^{s-2}},$$
$$\beta_{s-2} = \frac{\eta^{s-2}+\eta^{s-3}+\cdots+\eta+1}{\eta^{s-1}} = \eta - 1.$$

Let

$$\gamma_j = \begin{cases} \beta_j/\beta_{j+1}, & \text{if} \quad 0 \leqslant j \leqslant s-3, \\ \beta_j, & \text{if} \quad j = s-2. \end{cases}$$

Then

$$\gamma_{s-2} > \gamma_{s-3} > \cdots > \gamma_1 > \gamma_0, \tag{2.16}$$

since $\frac{x}{x+1}$ is an increasing function of $x$ for $x \geqslant 0$ and

$$\gamma_j = \frac{\eta^{j+1} + \eta^j + \cdots + \eta}{\eta^{j+1} + \eta^j + \cdots + \eta + 1} \quad (0 \leqslant j \leqslant s-2).$$

Let $x = \beta_{s-2}y$ and $g(y) = f(\beta_{s-2}y)$. Then

$$g(y) = \beta_{s-2}^{s-1}y^{s-1} + \beta_{s-2}^{s-1}y^{s-2} + \beta_{s-3}\beta_{s-2}^{s-3}y^{s-3} + \cdots + \beta_1\beta_{s-2}y + \beta_0.$$

From (2.16), we have

$$\beta_{s-2}^{s-1} > \beta_{s-3}\beta_{s-2}^{s-3} > \cdots > \beta_1\beta_{s-2} > \beta_0.$$

Hence it follows by Lemma 2.4 that the moduli of the roots of $g(y) = 0$ are all$\leqslant 1$, i.e., the moduli of the roots of $f(x) = 0$ are all$\leqslant \beta_{s-2} = \eta - 1$. (2.15) and so the lemma is proved.

**Lemma2.5** *$|\eta^{(2)}| = \eta^{-1}$ for $s = 2$ and $|\eta^{(2)}| = |\eta^{(3)}| = \eta^{-\frac{1}{2}}$ for $s = 3$.*

*Proof.* Obviously $|\eta^{(2)}| = \eta^{-1}$ for $s = 2$. For $s = 3$, since

$$x^3 - x^2 - x - 1 = (x - \eta)(x^2 + (\eta - 1)x + \eta^{-1}), \quad \eta > 1$$

and

$$\begin{aligned} (\eta - 1)^2 - 4\eta^{-1} &= \frac{\eta^3 - 2\eta^2 + \eta - 4}{\eta} = \frac{-\eta^2 + 2\eta - 3}{\eta} \\ &= -\frac{(\eta - 1)^2 + 2}{\eta} < 0, \end{aligned}$$

therefore $\eta^{(2)}$ and $\eta^{(3)}$ are conjugate complex numbers. Hence $|\eta^{(2)}| = |\eta^{(3)}| = \eta^{-\frac{1}{2}}$. The lemma is proved.

## 2.5 The roots of the equation $G(x) = 0$

Let $s \geqslant 2$. we denote the largest real root of the equation

$$G(x) = x^s - Lx^{s-1} - 1 = 0$$

by $\tau(= \tau^{(1)})$ and its other roots by $\tau^{(2)}, \cdots, \tau^{(s)}$, where $L$ is an integer$\geqslant 2$.

**Lemma2.6**

$$L < \tau < L + L^{-(s-1)} \tag{2.17}$$

*and*

$$(L + L^{-\frac{1}{s-1}})^{-\frac{1}{s-1}} < |\tau^{(i)}| < (L - (L-1)^{-\frac{1}{s-1}})^{-\frac{1}{s-1}},$$
$$2 \leqslant i \leqslant s. \tag{2.18}$$

*Proof.* (2.17) follows immediately by

$$G(L) = -1 < 0$$

and

$$G(L + L^{-(s-1)}) = L^{-(s-1)}(L + L^{-(s-1)})^{s-1} - 1 > 0.$$

Let $\chi$ be the real root of the equation

$$g(x) = x^s + Lx^{s-1} - 1 = 0$$

in the interval (0, 1). Since $g'(x) \neq 0$ in (0, 1), therefore $\chi$ is the only root of $g(x) = 0$ in (0, 1). It follows from

$$g(0) = -1 < 0$$

and

$$g(1) = L > 0$$

that $g(\chi - \delta) < 0$ for any $\delta$ satisfying $0 < \delta < \chi$. Hence on the circle in the complex plane $|x| = \chi - \delta$, we have

$$1 > |x^s + Lx^{s-1}|.$$

It follows by Rouché's theorem that 1 and $G(x)$ have the same number of zero in the circle $|x| < \chi - \delta$, i.e., $G(x)$ has no zero in the circle $|x| < \chi - \delta$. Let $\delta \to 0$. Then the moduli of roots of $G(x) = 0$ are all$\geqslant \chi$. Since

$$\begin{aligned} g((L + L^{-\frac{1}{s-1}})^{-\frac{1}{s-1}}) &= ((L + L^{-\frac{1}{s-1}})^{-\frac{1}{s-1}} + L)(L + L^{-\frac{1}{s-1}})^{-1} - 1 \\ &< (L^{-\frac{1}{s-1}} + L)(L + L^{-\frac{1}{s-1}})^{-1} - 1 = 0, \end{aligned}$$

we have the left hand side of (2.18)

$$|\tau^{(i)}| \geqslant \chi > (L + L^{-\frac{1}{s-1}})^{-\frac{1}{s-1}}, \quad 2 \leqslant i \leqslant s.$$

Suppose that $L > 2$. Let $\Omega$ be the only real root of the equation

$$h(x) = x^s - Lx^{s-1} + 1 = 0$$

in the interval (0, 1). Since

$$h(0) = 1 > 0$$

and

$$h(1) = -L + 2 < 0,$$

we have $h(\Omega + \delta) < 0$ for any $\delta$ satisfying $0 < \delta < 1 - \Omega$. Hence on the circle $|x| = \Omega + \delta$, we have

$$|Lx^{s-1}| > |x^s + 1|.$$

It follows by Rouché's theorem that $x^{s-1}$ and $G(x)$ have the same number of zeros in the circle $|x| < \Omega + \delta$, i.e., $G(x)$ has $s-1$ zeros in the circle $|x| < \Omega + \delta$. Let $\delta \to 0$. Then $G(x) = 0$ has $s-1$ roots in the circle $|x| \leqslant \Omega$. Since

$$\begin{aligned}
&h((L-(L-1)^{-\frac{1}{s-1}})^{-\frac{1}{s-1}}) \\
=&((L-(L-1)^{-\frac{1}{s-1}})^{-\frac{1}{s-1}} - L)(L-(L-1)^{-\frac{1}{s-1}})^{-1} + 1 \\
=&(L-(L-1)^{-\frac{1}{s-1}})^{-1}((L-(L-1)^{-\frac{1}{s-1}})^{-\frac{1}{s-1}} \\
&-(L-1)^{-\frac{1}{s-1}}) < 0,
\end{aligned}$$

we have

$$|\tau^{(i)}| \leqslant \Omega < (L-(L-1)^{-\frac{1}{s-1}})^{-\frac{1}{s-1}}, \quad 2 \leqslant i \leqslant s.$$

Suppose that $L = 2$. We proceed to prove that the equation

$$G(x) = x^s - 2x^{s-1} - 1 = 0$$

has $s-1$ roots with moduli <1. This is equivalent to proving that the equation

$$G^*(y) = y^s + 2y - 1 = 0$$

has $s-1$ roots with moduli >1. Let $\delta > 0$. Then it follows from Rouché's theorem that $G^*_\delta(y) = y^s + (2+\delta)y - 1$ and $y$ have the same number of zeros in the circle $|y| < 1$, i.e., $G^*_\delta(y) = 0$ has $s-1$ roots with moduli $\geqslant 1$. Let $\delta \to 0$. Then $G^*(y) = 0$ has also $s-1$ roots with moduli $\geqslant 1$. Since $G^*(y) = 0$ has no root satisfying $|y| = 1$, $G^*(y) = 0$ has $s-1$ roots with moduli $> 1$. The right hand side of (2.18) and also the lemma is proved.

**Lemma 2.7** *$|\tau^{(2)}| = \tau^{-1}$ for $s = 2$ and $|\tau^{(2)}| = |\tau^{(3)}| = \tau^{-\frac{1}{2}}$ for $s = 3$.*

*Proof.* Clearly $|\tau^{(2)}| = \tau^{-1}$ for $s = 2$. For $s = 3$, since

$$x^3 - Lx^2 - 1 = (x-\tau)(x^2 + (\tau - L)x + \tau^{-1}), \quad \tau > L$$

and

$$(\tau - L)^2 - 4\tau^{-1} = \frac{\tau^3 - 2L\tau^2 + L^2\tau - 4}{\tau} = \frac{-L\tau^2 + L^2\tau - 3}{\tau} < 0,$$

hence $\tau^{(2)}$ and $\tau^{(3)}$ are conjugate complex numbers and

$$|\tau^{(2)}| = |\tau^{(3)}| = \tau^{-\frac{1}{2}}.$$

The lemma is proved.

*Remark.* Minkowski proved that except when $Q(\alpha)$ is a real quadratic field or $Q(\alpha)$ is a cubic field and $\alpha^{(2)}$, $\alpha^{(3)}$ are conjugate complex numbers, the real algebraic number field $Q(\alpha)$ does not contain a unit $\eta$ such that

$$\eta > 1, \quad |\eta^{(2)}| = \cdots = |\eta^{(s)}|.$$

(Cf. H. Minhowshi [1], O. Perron [1]). However the above lemma shows that the absolute values of the conjugates of the unit $\tau$ are approximately equal if $L$ is sufficiently large.

## 2.6 The roots of the equation $H(x) = 0$

Let $s \geqslant 2$ and $A_1, \cdots, A_{s-1}$ be integers defined by the relation

$$A_1 = \binom{2s}{1}, \quad A_k = \binom{2s}{k} - A_1\binom{2s-2}{k-1} - \cdots - A_{k-1}\binom{2s-2k+2}{1},$$
$$s-1 \geqslant k > 1.$$

Let $r$ be the positive integer satisfying

$$s^2 > \frac{2}{r^{s-1}} + \frac{|A_1|}{r^{s-2}} + \cdots + \frac{|A_{s-2}|}{r}.$$

Further let $\omega (= \omega^{(1)})$ denote the largest real root of the equation

$$H(x) = x^s - s^2 r^{s-1} x^{s-1} + (-1)^{s-2} A_{s-2} r^{s-2} x^{s-2} + \cdots - A_1 r x - 1 = 0$$

and $\omega^{(2)}, \cdots, \omega^{(s)}$ its other roots.

**Lemma2.8**

$$\omega = s^2 r^{s-1} + O(r^{\frac{s}{2}-1}) \tag{2.19}$$

*and*

$$\omega^{(i)} = -\frac{1}{4\left(\sin\frac{\pi i}{s}\right)^2 r} + O(r^{-\frac{s}{2}-1}), \quad 1 \leqslant i \leqslant s-1, \tag{2.20}$$

*where the constant implied by the symbol "O" depends on s only.*

To prove Lemma 2.8, we shall need

**Lemma2.9**

$$(a+b)^{2s} - A_1 ab(a+b)^{2s-2} - A_2 a^2 b^2 (a+b)^{2s-4} - \cdots$$

$$-A_{s-2}a^{s-2}b^{s-2}(a+b)^4 + (-1)^{s-1}s^2a^{s-1}b^{s-1}(a+b)^2$$
$$-(a^s + (-1)^{s-1}b^s)^2 = 0. \tag{2.21}$$

*Proof.*
$$(a+b)^{2s} - A_1ab(a+b)^{2s-2}$$
is a symmetric function of $a, b$ without the term $ab^{2s-1}$.$(a+b)^{2s} - A_1ab(a+b)^{2s-2} - A_2a^2b^2(a+b)^{2s-4}$ is also a symmetric function of $a, b$ without the terms $ab^{2s-1}$ and $a^2b^{2s-2}$ and so on. Hence

$$(a+b)^{2s} - A_1ab(a+b)^{2s-2} - A_2a^2b^2(a+b)^{2s-4} - \cdots$$
$$-A_{s-1}a^{s-1}b^{s-1}(a+b)^2 - a^{2s} - b^{2s} - ka^sb^s = 0,$$

where $k$ is a constant. Let $a = -b$. Then

$$2b^{2s} + (-1)^sb^{2s}k = 0$$

and so

$$k = (-1)^{s-1}2.$$

Hence

$$(a+b)^{2s} - A_1ab(a+b)^{2s-2} - \cdots - A_{s-1}a^{s-1}b^{s-1}(a+b)^2$$
$$-(a^s + (-1)^{s-1}b^s)^2 = 0.$$

Divide the above formula by $(a+b)^2$ and then put $a = -b$. This yields that

$$(-1)^sA_{s-1}b^{2s-2} - s^2b^{2s-2} = 0,$$
$$A_{s-1} = (-1)^ss^2.$$

The lemma is proved.

The proof of Lemma 2.8. Put $a + b = y$, $ab = -r$ and $a^s + (-1)^{s-1}b^s = -1$ in (2.21). Then we have the equation

$$y^{2s} + A_1ry^{2s-2} - A_2r^2y^{2s-4} - \cdots + (-1)^{s-1}A_{s-2}r^{s-2}y^4$$
$$+s^2r^{s-1}y^2 - 1 = 0 \tag{2.22}$$

has $a + b$ as a solution, where $a, b$ satisfy

$$a = -\frac{r}{b}, \quad a^s + (-1)^{s-1}b^s = -1. \tag{2.23}$$

By (2.23), we have

$$b^{2s} + (-1)^{s-1}b^s - r^s = 0.$$

Denote $\zeta = e^{\pi i/s}$. Then

$$b = \left(\frac{(-1)^{s+l} + (1+4r^s)^{1/2}}{2}\right)^{1/s}\zeta^l$$

$$= \sqrt{r}\left(\frac{(-1)^{s+l}}{2r^{s/2}} + \left(1 + \frac{1}{4r^s}\right)^{1/2}\right)^{1/s} \zeta^l$$

$$= \sqrt{r}\left(1 + \frac{(-1)^{s+l}}{2r^{s/2}} + O(r^{-s})\right)^{1/s} \zeta^l$$

$$= \sqrt{r}\left(1 + \frac{(-1)^{s+l}}{2sr^{s/2}} + O(r^{-s})\right)^{1/s} \zeta^l,$$

$$1 \leqslant l \leqslant 2s.$$

Substituting into (2.23), we have

$$a = -\sqrt{r}\left(1 + \frac{(-1)^{s+l-1}}{2sr^{s/2}} + O(r^{-s})\right)\bar{\zeta}^l, \quad 1 \leqslant l \leqslant 2s$$

and so the roots of (2.22)

$$y = (\zeta^l - \zeta^l)\sqrt{r} + \frac{(-1)^{s+l}(\zeta^l + \bar{\zeta}^l)}{2sr^{\frac{s-1}{2}}} + O(r^{-s+\frac{1}{2}}), \quad 1 \leqslant l \leqslant 2s.$$

Setting the variable $z = y^2$ in equation (2.22), we know that the roots of the equation

$$z^s + A_1 r z^{s-1} - A_2 r^2 z^{s-2} - \cdots + (-1)^{s-1} A_{s-2} r^{s-2} z^2 + s^2 r^{s-1} z - 1 = 0 \qquad (2.24)$$

are

$$s^{-2} r^{-s+1} + O(r^{-\frac{3s}{2}+1}), \quad -4\left(\sin\frac{\pi i}{s}\right)^2 r + O(r^{-\frac{s}{2}+1}), \quad 1 \leqslant i \leqslant s-1.$$

Hence the roots of $H(x) = 0$ are (2.19) and (2.20) by substituting $x = y^{-1}$ in (2.24). The lemma is proved.

## 2.7 The irreducibility of a polynomial

Let

$$g(x) = x^s + a_{s-1}x^{s-1} + \cdots + a_1 x + a_0,$$

where $a_i (0 \leqslant i \leqslant s-1)$ are rational integers and $a_0 \neq 0$.

**Theorem2.4** *If*

$$|a_1| > |a_0^{s-1}| + |a_{s-1}a_0^{s-2}| + \cdots + |a_2 a_0| + 1, \qquad (2.25)$$

*then $g(x)$ is irreducible over $Q$.*

*Proof.* By (2.25), we have

$$|a_1 a_0| > |a_0^s| + |a_{s-1}a_0^{s-1}| + \cdots + |a_2 a_0^2| + |a_0|. \qquad (2.26)$$

It follows by Rouché's theorem that $g(x)$ and $x$ have the same number of zeros in the circle $|x| < |a_0|$, i.e., $g(x)$ has only one zero $\vartheta$ in the circle $|x| < a_0$. By (2.26), the equation $g(x) = 0$ has no root with modulus $|a_0|$.

If $g(x) = u(x)v(x)$, where $u(x)$ and $v(x)$ are polynomials with integral coefficients and with degrees $\geqslant 1$, and if $u(\vartheta) = 0$, then the moduli of the roots of $v(x) = 0$ are all $> |a_0|$. Hence

$$|a_0| = |g(0)| = |u(0)v(0)| \geqslant |v(0)| > |a_0|$$

which leads to a contradiction. Thus we have the theorem.

**Theorem2.5** *If $g(x) = 0$ has only a root $\vartheta$ with modulus $\geqslant 1$, then $g(x)$ is irreducible over $Q$.*

*Proof* If $g(x) = u(x)v(x)$, where $u(x)$ and $v(x)$ are polynomials with integral coefficients and with degrees $\geqslant 1$, and if $u(\vartheta) = 0$, then the moduli of the roots of $v(x)$ are all $< 1$. Hence $|v(0)| < 1$ which leads to a contradiction. The theorem follows.

**Theorem2.6** *If*

$$|a_{s-1}| > |a_{s-2}| + \cdots + |a_1| + |a_0| + 1, \tag{2.27}$$

*then $g(x)$ is irreducible over $Q$.*

*Proof.* It follows by (2.27) and Rouché's theorem that $x^{s-1}$ and $g(x)$ have the same number of zeros in the circle $|x| < 1$. Hence $g(x)$ has only one zero $\vartheta$ with modulus $\geqslant 1$ and thus the theorem follows by Theorem 2.5.

**Theorem2.7** *$\eta$, $\tau$, $\omega$ are all* PV *numbers of degree $s$.*

*Proof.* Since

$$\eta > 1, \quad |\eta^{(i)}| < 1, \quad 2 \leqslant i \leqslant s$$

by Lemma 2.3, $F(x)$ is irreducible over $Q$ by Theorem 2.5 and so $\eta$ is a PV number of degree $s$. Similarly, we may prove that $\tau$ is a PV number of degree $s$ too. By Theorem 2.6 and its proof, we know that $H(x)$ is irreducible over $Q$ and $\omega$ is a PV number of degree $s$. The theorem is proved.

## 2.8 The rational approximations of $\eta$,$\tau$,$\omega$

1. Let $F_n (= F_{s,n}) (n = 0, 1, \cdots)$ be a sequence of integers defined by the recurrence relation

$$F_0 = F_1 = \cdots = F_{s-2} = 0, \quad F_{s-1} = 1,$$
$$F_{n+s} = F_{n+s-1} + F_{n+s-2} + \cdots + F_{n+1} + F_n, \quad n \geqslant 0.$$

As usual, $(F_n)$ is called the generalized Fibonacci sequence of dimension $s$. Let

$$\rho = -\frac{\ln|\eta^{(s)}|}{\ln\eta}.$$

Then

$$\rho \geqslant -\frac{\ln(\eta-1)}{\ln\eta} \geqslant -\frac{\ln(1-2^{-s})}{\ln 2} \geqslant \frac{1}{2^s\ln 2}+\frac{1}{2^{2s+1}}.$$

Take a basis of $Q(\eta)$

$$\omega_1=1,\quad \omega_2=\eta-1,\cdots,\quad \omega_s=\eta^{s-1}-\eta^{s-2}-\cdots-\eta-1.$$

Set

$$F_n(j)=F_{n+j-1}-F_{n+j-2}-\cdots-F_{n+1}-F_n,\quad 2\leqslant j\leqslant s$$

Then we can derive from Theorem 2.3 the following

**Theorem 2.8** *For $n\geqslant s$, we have*

$$\left|\frac{F_n(j)}{F_n}-\omega_j\right| \leqslant c(\eta)F_n^{-1-\frac{1}{2^s\ln 2}-\frac{1}{2^{2s+1}}},\quad 2\leqslant j\leqslant s. \tag{2.28}$$

For the cases $s=2$ and $s=3$, if we use Lemma 2.5 to replace Lemma 2.3, then we have

**Theorem2.9** *The right hand side of* (2.28) *may be replaced by $c(\eta)F_n^{-2}$ and $c(\eta)F_n^{-\frac{3}{2}}$ for the cases $s=2$ and $s=3$ respectively.*

2. Let $G_n(=G_{s,n})(n=0,1,\cdots)$ be the sequence of integers defined by the recurrence relation

$$G_0=G_1=\cdots=G_{s-2}=0,\quad G_{s-1}=1, G_{n+s}=LG_{n+s-1}+G_n,\quad n\geqslant 0.$$

Let

$$\rho=-\frac{\ln|\tau^{(s)}|}{\ln\tau}.$$

Since

$$\begin{aligned}\ln(L-(L-1)^{-\frac{1}{s-1}}) &= \ln L+\ln(1-L^{-1}(L-1)^{-\frac{1}{s-1}})\\ &\geqslant \ln L-L^{-1}(L-1)^{-\frac{1}{s-1}}-L^{-2}(L-1)^{-\frac{2}{s-1}}\\ &\geqslant \ln L-L^{-1}-L^{-2}\end{aligned}$$

and

$$\ln(L+L^{-(s-1)})=\ln L+\ln(1+L^{-s})\leqslant \ln L+L^{-s},$$

therefore

$$\begin{aligned}\rho &\geqslant \frac{\ln L-L^{-1}-L^{-2}}{(s-1)(\ln L+L^{-s})}\\ &\geqslant \frac{1}{s-1}\left(1-\frac{1}{L\ln L}-\frac{1}{L^2\ln L}-\frac{1}{L^s\ln L}+\frac{1}{L^{s+1}(\ln L)^2}\right)\end{aligned}$$

$$\geqslant \frac{1}{s-1}\left(1-\frac{2}{L\ln L}+\frac{1}{L^{s+3}}\right)$$

by Lemma2.6. Hence we can derive from Theorem 2.2 the following

**Theorem2.10** *for $n \geqslant s$, we have*

$$\left|\frac{G_{n+j}}{G_n}-\tau^j\right| \leqslant c(\tau)G_n^{-1-\frac{1}{s-1}+\frac{2}{(s-1)L\ln L}-\frac{1}{(s-1)L^{s+3}}}, \quad 1\leqslant j\leqslant s-1. \tag{2.29}$$

For the cases $s=2$ and $s=3$, if we use Lemma 2.7 to replace Lemma 2.6, then we have

**Theorem2.11** *The right hand side of* (2.29) *may be replaced by $c(\tau)G_n^{-2}$ and $c(\tau)G_n^{-\frac{3}{2}}$ for the cases $s=2$ and $s=3$ respectively.*

3. Let $H_n(=H_{s,n})(n=0,1,\cdots)$ be the sequence of integers defined by the recurrance relation

$$\begin{gathered} H_0=H_1=\cdots=H_{s-2}=0, \quad H_{s-1}=1, \\ H_{n+s}=s^2r^{s-1}H_{n+s-1}+(-1)^{s-1}A_{s-2}r^{s-2}H_{n+s-2}+\cdots \\ +A_1rH_{n+1}+H_n, \quad n\geqslant 0. \end{gathered}$$

Let

$$\rho=-\frac{\ln|\omega^{(s)}|}{\ln\omega}.$$

Then by Lemma 2.8, we have

$$\begin{aligned} \rho &= \frac{\ln\left(4r\left(\sin\frac{\pi}{s}\right)^2\right)+O(r^{-s/2})}{\ln(s^2r^{s-1})+O(r^{-s/2})} \\ &= \frac{\ln r+2\ln\left(2\sin\frac{\pi}{s}\right)+O(r^{-s/2})}{(s-1)\ln r+2\ln s+O(r^{-s/2})} \\ &= \frac{1}{s-1}+\frac{c_1}{\ln r}+\frac{c_2}{(\ln r)^2}+O\left(\frac{1}{(\ln r)^3}\right). \end{aligned} \tag{2.30}$$

where

$$\begin{aligned} c_1 &= \frac{2\ln\left(2\sin\frac{\pi}{s}\right)}{s-1}-\frac{2\ln s}{(s-1)^2}, \\ c_2 &= -\frac{4\ln s\ln\left(2\sin\frac{\pi}{s}\right)}{(s-1)^2}+\frac{4(\ln s)^2}{(s-1)^3}. \end{aligned} \tag{2.31}$$

Clearly $H_n$ increases with $n$. We can derive from Theorem 2.2 the following

**Theorem2.12** *For $n\geqslant s$, we have*

$$\left|\frac{H_{n+j}}{H_n}-\omega^j\right| \leqslant c(\omega)H_n^{-1-\rho}, \quad 1\leqslant j\leqslant s-1.$$

*Where* $\rho$ *is defined by* (2.30) *and* (2.31).

*Remarks.* 1. Concerning the generalization of Fibonacci sequence, except those given here and in §1.3, Raney [1] also gave a generalization and his result may be obtained from the results of §2.1 and §2.2 (Cf. G. N. Raney [1]).

2. Although the errors in rational approximations of $\tau$ and $\omega$ are better, the sequences of $G_n$ and $H_n$ increase too fast as $n$ increases and so they are not as convenient in practical uses as compared with the sequence $F_n$.

**Notes**

The definition of PV number was first introduced by C. Pisot[1] and T. Vijayaraghavan[1] (Cf. J. W. S. Cassels[1]).

Lemma2.8: Cf. Hua Loo Keng [1].

Theorem 2.4 is due to Xie Ting Fan and Pei Ding Yi[1] which improves a theorem of O. Perron [1] and also a theorem of L. Bernstein[1].

The other results: Cf. Hua Loo Keng and Wang Yuan [6,7,8].

# Chapter 3

# Uniform Distribution

## 3.1 Uniform distribution

We use $G_s$ to denote the $s$-dimensional unit cube

$$0 \leqslant x_i \leqslant 1, \quad 1 \leqslant i \leqslant s.$$

Let $n$ be a positive integer and

$$P_n(k) = (x_1^{(n)}(k), \cdots, x_s^{(n)}(k)), \quad 1 \leqslant k \leqslant n$$

be a set of points in $G_s$, where

$$0 \leqslant x_i^{(n)}(k) < 1, \quad 1 \leqslant i \leqslant s.$$

For any $\gamma = (\gamma_1, \cdots, \gamma_s) \in G_s$, let $N_n(\gamma) = N_n(\gamma_1, \cdots, \gamma_s)$ denote the number of points of $P_n(k)(1 \leqslant k \leqslant n)$ satisfying the inequalities

$$0 \leqslant x_i^{(n)}(k) < \gamma_i, \quad 1 \leqslant i \leqslant s.$$

Then

$$\sup_{\gamma \in G_s} \left| \frac{N_n(\gamma)}{n} - |\gamma| \right| = D(n), \quad |\gamma| = \gamma_1 \cdots \gamma_s$$

is called the discrepancy of the set of points $P_n(k)(1 \leqslant k \leqslant n)$. Let $n_1 < n_2 < \cdots$ be a sequence of positive integers. Let

$$P_{nl}(k) = (x_1^{(n_l)}(k), \cdots, x_s^{(n_l)}(k)), \quad 1 \leqslant k \leqslant n_l$$

be a set of points in $G_s$ with discrepancy $D(n_l)$. If $D(n_l) = o(1)$, then the sequence of sets $P_{n_l}(k)(n_1 < n_2 < \cdots)$ is said to be uniformly distributed with discrepancy $D(n)$. In case $n_l = l, x_1^{(l)}(k) = x_1(k), \cdots, x_s^{(l)}(k) = x_s(k)(k = 1, 2, \cdots)$, the sequence $P(k) = (x_1(k), \cdots, x_s(k))(k = 1, 2, \cdots)$ will be called uniformly distributed in $G_s$.

## 3.2 Vinogradov's lemma

**Lemma3.1** *Let $r$ be a positive integer. Let $\alpha, \beta, \Delta$ be real numbers satisfying*

$$0 < \Delta < \frac{1}{2}, \quad \Delta \leqslant \beta - \alpha \leqslant 1 - \Delta.$$

*Then there exists periodic function $\Psi(x)$ with period 1 such that*

1) $\Psi(x) = 1$, *if* $\alpha + \frac{1}{2}\Delta \leqslant x \leqslant \beta - \frac{1}{2}\Delta$,

2) $0 \leqslant \Psi(x) \leqslant 1$, *if* $\alpha - \frac{1}{2}\Delta \leqslant x \leqslant \alpha + \frac{1}{2}\Delta$ *and*

$\beta - \frac{1}{2}\Delta \leqslant x \leqslant \beta + \frac{1}{2}\Delta$,

3) $\Psi(x) = 0$, *if* $\beta + \frac{1}{2}\Delta \leqslant x \leqslant 1 + \alpha - \frac{1}{2}\Delta$,

4) $\Psi(x)$ *has a Fourier expansion*

$$\Psi(x) = \beta - \alpha + \sum{}' C(m) e^{2\pi i m x},$$

*where* $\sum'$ *denotes a sum with $m = 0$ deleted and*

$$|C(m)| \leqslant \min\left(\beta - \alpha, \frac{1}{\pi|m|}, \left(\frac{1}{\pi|m|}\right)^{r+1} \left(\frac{r}{\Delta}\right)^r\right).$$

*Proof.* Define a periodic function with period 1

$$\Psi_0(x) = \begin{cases} 1, & \text{if} \quad \alpha < x < \beta, \\ \frac{1}{2}, & \text{if} \quad x = \alpha \quad \text{of} \quad x = \beta, \\ 0 & \text{if} \quad \beta < x < 1 + \alpha. \end{cases}$$

Then $\Psi_0(x)$ has the Fourier expansion

$$\Psi_0(x) = C_0^{(0)} + \sum{}' C_m^{(0)} e^{2\pi i m x},$$

where

$$\begin{aligned} C_0^{(0)} &= \int_0^1 \Psi_0(x) dx = \beta - \alpha, \\ C_m^{(0)} &= \int_0^1 \Psi_0(x) e^{-2\pi i m x} dx = \int_\alpha^\beta e^{-2\pi i m x} dx \\ &= \frac{e^{-2\pi i m \alpha} - e^{-2\pi i m \beta}}{2\pi i m} (m \neq 0), \end{aligned}$$

and so

$$|C_m^{(0)}| \leqslant \min\left(\beta - \alpha, \frac{1}{\pi|m|}\right).$$

Let $\Delta = 2r\delta$ and

$$\Psi_\rho(x) = \frac{1}{2\delta}\int_{-\delta}^{\delta} \Psi_{\rho-1}(x+z)dz,$$

where $\rho = 1, 2, \cdots, r$. Then we may prove by induction that

1)′ $\Psi_\rho(x) = 1$, if $\alpha + \rho\delta < x < \beta - \rho\delta$,

2)′ $0 \leqslant \Psi_\rho(x) \leqslant 1$, if $\alpha - \rho\delta \leqslant x \leqslant \alpha + \rho\delta$ and $\beta - \rho\delta \leqslant x \leqslant \beta + \rho\delta$,

3)′ $\Psi_\rho(x) = 0$, if $\beta + \rho\delta < x < 1 + \alpha - \rho\delta$,

4)′ $\Psi_\rho(x)$ has a Fourier expansion

$$\Psi_\rho(x) = C_0^{(\rho)} + \sum{}' C_m^{(\rho)} e^{2\pi i m x},$$

where

$$C_0^{(\rho)} = \beta - \alpha,$$
$$|C_m^{(\rho)}| \leqslant \min\left(\beta - \alpha, \frac{1}{\pi|m|}, \left(\frac{1}{\pi|m|}\right)^{\rho+1}\left(\frac{r}{\Delta}\right)^{\rho}\right).$$

In fact, suppose that $\Psi_{\rho-1}(x)$ satisfies 1)′ − 4)′. Then $\Psi_\rho(x)$ satisfies 1)′ − 3)′ obviously. Now we prove that $\Psi_\rho(x)$ satisfies 4)′ as follows:

$$\begin{aligned} C_0^{(\rho)} &= \int_0^1 \Psi_\rho(x)dx = \frac{1}{2\delta}\int_{-\delta}^{\delta} dz \int_0^1 \Psi_{\rho-1}(x+z)dx \\ &= C_0^{(\rho-1)} = \beta - \alpha, \end{aligned}$$

$$\begin{aligned} C_m^{(\rho)} &= \int_0^1 \Psi_\rho(x)e^{-2\pi i m x}dx = \frac{1}{2\delta}\int_0^1 e^{-2\pi i m x}dx \int_{-\delta}^{\delta} \Psi_{\rho-1}(x+z)dz \\ &= \frac{C_m^{(\rho-1)}}{2\delta}\int_{-\delta}^{\delta} e^{2\pi i m z}dz = C_m^{(\rho-1)}\left(\frac{e^{2\pi i m\delta} - e^{-2\pi i m\delta}}{4\pi i m\delta}\right) \\ &= C_m^{(0)}\left(\frac{e^{2\pi i m\delta} - e^{-2\pi i m\delta}}{4\pi i m\delta}\right)^{\rho}. \end{aligned}$$

Hence

$$|C_m^{(\rho)}| \leqslant \min\left(\beta - \alpha, \frac{1}{\pi|m|}, \left(\frac{1}{\pi|m|}\right)^{\rho+1}\left(\frac{r}{\Delta}\right)^{\rho}\right).$$

Take $\Psi(x) = \Psi_r(x)$. Then we have the lemma.

## 3.3 The exponential sum and the discrepancy

We use the notations $\bar{x} = \max(1, |x|)$, $\|\gamma\| = \bar{\gamma}_1 \cdots \bar{\gamma}_s$ and $(\alpha, \beta) = \sum_{i=1}^{s} \alpha_i \beta_i$ the scalar product of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

**Theorem3.1** *Let $r, h$ be positive integers such that $h > r/\eta$, where $\eta$ satisfies $0 < \eta < 1/6$. Then for any $\gamma \in G_s$, we have*

$$\left|\frac{1}{n} N_n(\gamma) - |\gamma|\right| < D(n)$$

*where*

$$D(n) = \sum_{\|m\| \leqslant h}{}' \frac{1}{\|\pi m\|} \left| \frac{1}{n} \sum_{k=1}^{n} e^{2\pi i(\mathbf{m}, P_n(k))} \right| + (5s+6)\eta + \frac{s 2^s r^{r-1}}{\pi^{s+r} \eta^r h^r} (\ln 64h)^{s-1},$$

*is which $\sum'$ denotes a sum with $\mathbf{m} = \mathbf{0} = (0, \cdots, 0)$ deleted.*

**Lemma3.2** *Let $n$, $l$ be integers $> 1$. Then*

$$\sum_{m=1}^{n} \frac{1}{m} < 1 + \ln n$$

*and*

$$\sum_{m=n+1}^{\infty} \frac{1}{m^l} < \frac{1}{l-1} n^{-l+1}.$$

*Proof.* Since

$$\frac{1}{m} < \int_{m-1}^{m} \frac{dt}{t},$$

hence

$$\sum_{m=1}^{n} \frac{1}{m} < 1 + \int_1^2 \frac{dt}{t} + \cdots + \int_{n-1}^{n} \frac{dt}{t} = 1 + \int_1^n \frac{dt}{t} = 1 + \ln n.$$

Similarly

$$\sum_{m=n+1}^{\infty} \frac{1}{m^l} < \int_n^{\infty} \frac{dt}{t^l} = \frac{n^{-l+1}}{l-1}.$$

The lemma is proved.

The proof of Theorem 1. First Suppose that $\gamma \in G_s$, where $\gamma_i' s$ satisfy

$$3\eta \leqslant \gamma_i \leqslant 1 - 3\eta, \quad 1 \leqslant i \leqslant s.$$

Let

$$G_x(y) = \begin{cases} 1 & \text{if} \quad 0 \leqslant y < x, \\ 0, & \text{if} \quad x \leqslant y < 1. \end{cases}$$

Then

$$\frac{1}{n} N_n(\gamma) = \frac{1}{n} \sum_{k=1}^{n} \sum_{\substack{x_\nu^{(n)}(k) < \gamma_\nu \\ 1 \leqslant \nu \leqslant s}} 1 = \frac{1}{n} \sum_{k=1}^{n} \prod_{\nu=1}^{s} G_{\gamma_\nu}(x_\nu^{(n)}(k)). \tag{3.1}$$

For $3\eta \leqslant x \leqslant 1 - 3\eta$, we construct two auxiliary functions $G_x^{(1)}(y)$ and $G_x^{(2)}(y)$, where $G_x^{(1)}(y)$ satisfies

1) $G_x^{(1)}(y) = 1$, if $-\eta \leqslant y \leqslant x$,
2) $0 \leqslant G_x^{(1)}(y) \leqslant 1$, if $-2\eta \leqslant y \leqslant -\eta$ and $x \leqslant y \leqslant x + \eta$,
3) $G_x^{(1)}(y) = 0$, if $x + \eta \leqslant y \leqslant 1 - 2\eta$,
4) $G_x^{(1)}(y)$ has the Fourier expansion

$$G_x^{(1)}(y) = x + 2\eta + {\sum}' C_1(m) e^{2\pi i m y},$$

in which

$$|C_1(m)| \leqslant \min\left(x + 2\eta, \frac{1}{\pi|m|}, \frac{r^r}{\pi^{r+1}\eta^r |m|^{r+1}}\right),$$

and where $G_x^{(2)}(y)$ satisfies

1)′ $G_x^{(2)}(y) = 1$, if $2\eta \leqslant y \leqslant x - \eta$,
2)′ $0 \leqslant G_x^{(2)}(y) \leqslant 1$, if $\eta \leqslant y \leqslant 2\eta$ and $x - \eta \leqslant y \leqslant x$,
3)′ $G_x^{(2)}(y) = 0$, if $x \leqslant y \leqslant 1 + \eta$,
4)′ $G_x^{(2)}(y)$ has the Fourier expansion

$$G_x^{(2)}(y) = x - 2\eta + {\sum}' C_2(m) e^{2\pi i m y},$$

in which

$$|C_2(m)| \leqslant \min\left(x - 2\eta, \frac{1}{\pi|m|}, \frac{r^r}{\pi^{r+1}\eta^r |m|^{r+1}}\right).$$

Then

$$G_x^{(2)}(y) \leqslant G_x(y) \leqslant G_x^{(1)}(y). \tag{3.2}$$

By Lemma 3.2

$$\begin{aligned} G_x^{(1)}(y) &= \sum_{|m| \leqslant h} C_1(m) e^{2\pi i m y} + \theta \sum_{|m| > h} \frac{r^r}{\pi^{r+1}\eta^r |m|^{r+1}} \\ &= \sum_{|m| \leqslant h} C_1(m) e^{2\pi i m y} + \frac{2\theta r^{r-1}}{\pi^{r+1}\eta^r h^r}, \end{aligned} \tag{3.3}$$

where $C_1^{(0)} = x + 2\eta$ and we use $\theta$ to denote a number satisfying $0 \leqslant |\theta| \leqslant 1$ but not always with the same value. From (3.1), (3.2), (3.3) and Lemma 3.2, we have

$$\begin{aligned} \frac{1}{n} N_n(\gamma) &\leqslant \frac{1}{n} \sum_{k=1}^{n} \prod_{\gamma=1}^{s} G_{\gamma_\nu}^{(1)}(x_\nu^{(n)}(k)) \\ &= \frac{1}{n} \sum_{k=1}^{n} \prod_{\gamma=1}^{s} \left( \sum_{|m| \leqslant h} C_1(m) e^{2\pi i m x_\nu^{(n)}(k)} + \frac{2\vartheta r^{r-1}}{\pi^{r+1} \eta^r h^r} \right) \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} 1 + \sum_{|m| \leqslant h} |C_1(m)| &\leqslant 2 + \frac{2}{\pi} \sum_{m=1}^{h} \frac{1}{m} \\ &\leqslant 2 + \frac{2}{\pi} + \frac{2}{\pi} \ln h < \frac{2}{\pi} \ln 64h. \end{aligned} \tag{3.5}$$

Since

$$(\gamma_1 + 2\eta) \cdots (\gamma_s + 2\eta) \leqslant \gamma_1(\gamma_2 + 2\eta) \cdots (\gamma_s + 2\eta) + 2\eta \leqslant \cdots \leqslant |\gamma| + 2s\eta, \tag{3.6}$$

so from (3.4), (3.5), (3.6), we have

$$\begin{aligned} \frac{1}{n} N_n(\gamma) - |\gamma| &\leqslant \left| \sum_{|m_i| \leqslant h}{}' C_1(m) \cdots C_1(m_s) \frac{1}{n} \sum_{k=1}^{n} e^{2\pi i (\mathbf{m}, P_n(k))} \right| \\ &\quad + 2s\eta + \frac{2sr^{r-1}}{\pi^{r+1} \eta^r h^r} \left( \frac{2}{\pi} \right)^{s-1} (\ln 64h)^{s-1}. \end{aligned}$$

Using $G_x^{(2)}(y)$ to instead of $G_x^{(1)}(y)$, we obtain a similar lower estimate for $\frac{1}{n} N_n(\boldsymbol{\gamma}) - |\boldsymbol{\gamma}|$. Hence

$$\begin{aligned} \left| \frac{1}{n} N_n(\boldsymbol{\gamma}) - |\boldsymbol{\gamma}| \right| &\leqslant \sum_{|m_i| \leqslant h}{}' \frac{1}{\|\pi \mathbf{m}\|} \left| \frac{1}{n} \sum_{k=1}^{n} e^{2\pi i (\mathbf{m} P_n(k))} \right| \\ &\quad + 2s\eta + \frac{s 2^s r^{r-1}}{\pi^{r+s} \eta^r h^r} (\ln 64h)^{s-1} = \Phi(\text{say}). \end{aligned}$$

Next, suppose that there are $t$ components of $\boldsymbol{\gamma}$ satisfying $\gamma_i < 3\eta$ and the rest satisfying $3\eta \leqslant \gamma_i \leqslant 1 - 3\eta$. Define $\boldsymbol{\gamma}' = (\gamma_1', \cdots, \gamma_s')$, where $\gamma_i' = 3\eta$ if $\gamma_i < 3\eta$, and $\gamma_i' = \gamma_i$ otherwise. Then $N_n(\boldsymbol{\gamma}') \geqslant N_n(\boldsymbol{\gamma})$ and

$$\left| \frac{N_n(\boldsymbol{\gamma})}{n} - |\boldsymbol{\gamma}| \right| \leqslant \frac{N_n(\boldsymbol{\gamma}')}{n} + |\boldsymbol{\gamma}| \leqslant \left| \frac{N_n(\boldsymbol{\gamma}')}{n} - |\boldsymbol{\gamma}'| \right| + |\boldsymbol{\gamma}'| + |\boldsymbol{\gamma}| \leqslant \Phi + 6\eta. \tag{3.7}$$

Finally, suppose that there are $t$ components of $\boldsymbol{\gamma}$ equal to 1 and the rest are not greater than $1-3\eta$. Then the problem is reduced to the $s-t$ dimensional case. Hence we have (3.7) too. For any $\boldsymbol{\gamma} \in G_s$, define $\boldsymbol{\gamma}' = (\gamma_1', \cdots, \gamma_s')$ and $\boldsymbol{\gamma}'' = (\gamma_1'', \cdots, \gamma_s'')$ as follows:

$$\gamma_i' = \begin{cases} 1-3\eta, & \text{if} \quad \gamma_i > 1-3\eta, \\ \gamma_i, & \text{if} \quad \gamma_i \leqslant 1-3\eta \end{cases}$$

and

$$\gamma_i'' = \begin{cases} 1 & \text{if} \quad \gamma_i > 1-3\eta, \\ \gamma_i, & \text{if} \quad \gamma_i \leqslant 1-3\eta. \end{cases}$$

Hence

$$\left|\frac{N_n(\boldsymbol{\gamma})}{n} - |\boldsymbol{\gamma}|\right| \leqslant \max\left(\left|\frac{N_n(\boldsymbol{\gamma}')}{n} - |\boldsymbol{\gamma}|\right|, \left|\frac{N_n(\boldsymbol{\gamma}'')}{n} - |\boldsymbol{\gamma}|\right|\right) \leqslant \Phi + 6\eta + 3s\eta.$$

The theorem follows.

## 3.4 The number of solutions to the congruence

We shall always use $M$ to denote a number $\geqslant 1$, $n$ an integer $\geqslant 2$ and $\mathbf{a} = (a_1, \cdots, a_s)$ an integral vector.

**Lemma3.3** *Suppose that the congruence*

$$(\mathbf{a}, \mathbf{m}) = \sum_{i=1}^{s} a_i m_i \equiv 0(\text{mod} n) \tag{3.8}$$

*has no solution in the domain*

$$\|\boldsymbol{m}\| \leqslant M, \quad \mathbf{m} \neq \mathbf{0}.$$

*Then the number of solutions $T_{l,M}$ of (3.8) in the domain*

$$\|\mathbf{m}\| < lM \tag{3.9}$$

*satisfies*

$$T_{l,M} \leqslant c(\varepsilon)^s l^{l+\varepsilon} M^{\varepsilon},$$

*where $l$ is a positive integer.*

We use $P_{s,M}$ to denote the $s$-dimensional parallelepiped with edges parallel to coordinate axes and with volume $\leqslant M$. To prove Lemma3.3, we shall need

**Lemma3.4** *The domain (3.9) can be covered by at most $c(\varepsilon)^s l^{1+\varepsilon} M^{\varepsilon}$ parallelepipeds of the type $P_{s,M}$.*

*Proof.* Take

$$c(\varepsilon) = 2^{2+\varepsilon} \sum_{j=0}^{\infty} (\bar{j}^{-(1+\varepsilon)} + 2^{-\varepsilon j}).$$

For $s = 1$, the domain (3.9) is an open interval $(-lM, lM)$, hence it can be covered by

$$\frac{2lM}{M} = 2l$$

intervals of the type $[c, c + M]$, where $c$ is a constant. The lemma holds.

Suppose that $k$ is a positive integer and the lemma holds for $s = 1, 2, \cdots, k$. Now we proceed to prove that the lemma holds for $s = k + 1$ too.

Divide the domain

$$\bar{m}_1 \cdots \bar{m}_{k+1} < lM$$

into $2[\log_2 M] + 3$ parts

$$m_{k+1} = j, \quad |j| \leqslant l, \tag{3.10}$$

and

$$2^i l < |m_{k+1}| \leqslant 2^{i+1} l, \quad 0 \leqslant i \leqslant [\log_2 M]. \tag{3.11}$$

Suppose that $m_{k+1} = j$. Then

$$\bar{m}_1 \cdots \bar{m}_k < \frac{lM}{\bar{j}} < \left(\left[\frac{l}{\bar{j}}\right] + 1\right) M,$$

so it can be covered by at most

$$Q = c(\varepsilon)^k \left(\left[\frac{l}{\bar{j}}\right] + 1\right)^{1+\varepsilon} M^\varepsilon$$

parallelepipeds of the type $P_{k,M}$ from the induction hypothesis. Using $P_{k,M}$ as basis and 1 as height, we obtain a $P_{k+1,M}$. Hence the subdomain $m_{k+1} = j$ can be covered by at most $Q$ parallelepipeds of the type $P_{k+1,M}$ and so the domain (3.10) can be covered by at most

$$2\sum_{j=0}^{l} c(\varepsilon)^k \left(\left[\frac{l}{\bar{j}}\right] + 1\right)^{1+\varepsilon} M^\varepsilon \tag{3.12}$$

parallelepipeds of the type $P_{k+1,M}$.

Consider the subdomain of (3.11)

$$2^i l < m_{k+1} \leqslant 2^{i+1} l. \tag{3.13}$$

Then we have

$$\bar{m}_1 \cdots \bar{m}_k < \frac{M}{2^i}$$

for $m_{k+1} = 2^i l + 1$ and it can be covered by at most

$$c(\varepsilon)^k \left(\frac{M}{2^i}\right)^\varepsilon$$

parallelepipeds of the type $P_{k,\frac{M}{2^i}}$ from the induction hypothesis. Since we may obtain a $P_{k+1,M}$ by the use of $P_{k,\frac{M}{2^i}}$ as basis and $2^i$ as height and $2^i l + 2^i l + 1 > 2^{i+1} l$, the domain (3.13) can be covered by at most

$$c(\varepsilon)^k l \left(\frac{M}{2^i}\right)^\varepsilon$$

parallelepipeds of the type $P_{k+1,M}$ and so the domain (3.11) is covered by at most

$$2c(\varepsilon)^k l \sum_{i=0}^{[\log_2 M]} \left(\frac{M}{2^i}\right)^\varepsilon \tag{3.14}$$

parallelepipeds of the type $P_{k+1,M}$.

It follows by (3.13) and (3.14) that the domain (3.9) can be covered by at most

$$c(\varepsilon)^k l^{1+\varepsilon} M^\varepsilon \sum_{j=0}^{\infty} \left(\frac{2^{2+\varepsilon}}{j^{1+\varepsilon}} + \frac{2}{2^{\varepsilon j}}\right) \leqslant c(\varepsilon)^{k+1} l^{1+\varepsilon} M^\varepsilon$$

parallelepipeds of the type $P_{k+1,M}$. Hence the lemma follows by induction.

The proof of Lemma3.3. By Lemma 3.4, it is sufficient to prove that the congruence (3.8) has at most 1 solution in any parallelepiped of the type $P_{s,M}$. Suppose that (3.8) has two solutions $\mathbf{m}'$ and $\mathbf{m}''$ in a $P_{s,M}$, where $\mathbf{m}' \neq \mathbf{m}''$. Let $\mathbf{m} = \mathbf{m}' - \mathbf{m}''$. Then $\|\mathbf{m}\| \leqslant M$, $\mathbf{m} \neq \mathbf{0}$ and

$$(\mathbf{a}, \mathbf{m}) = (\mathbf{a}, \mathbf{m}') - (\mathbf{a}, \mathbf{m}'') \equiv 0 (\bmod n)$$

which leads to a contradiction. The lemma follows.

**Lemma3.5** *Under the assumption of Lemma* 3.3,

$$T_{l,M} \leqslant c(s) l (\ln 3lM)^{s-1}.$$

*Proof.* We may prove easily by induction that the conclusion $c(\varepsilon)^s l^{1+\varepsilon} M^\varepsilon$ in Lemma 3.4 can be replaced by $c(s) l (\ln 3lM)^{s-1}$ and so the lemma follows.

## 3.5 The solutions of the congruence and the discrepancy

**Theorem3.2** *Under the assumption of Lemma* 3.3, *the set*

$$\left(\left\{\frac{a_1 k}{n}\right\}, \cdots, \left\{\frac{a_s k}{n}\right\}\right), \quad 1 \leqslant k \leqslant n$$

*has discrepancy*

$$D(n) \leqslant c(s) M^{-1} (\ln 3M)^s.$$

**Lemma3.6**

$$\frac{1}{n}\sum_{k=1}^{n} e^{2\pi imk/n} = \begin{cases} 1, & \text{if} \quad n|m, \\ 0, & \text{if} \quad n\nmid m. \end{cases}$$

(Cf. Hua Loo Keng [2], Chap. 7).

The proof of Theorem 3.2. Let $h \geqslant 7$. Then by Lemmas 3.5 and 3.6,

$$\begin{aligned}
&\sum_{|m_i|\leqslant h}{}' \frac{1}{\|\pi \mathbf{m}\|}\left|\frac{1}{n}\sum_{k=1}^{n} e^{2\pi i(\mathbf{a},\mathbf{m})k/n}\right| \leqslant \sum_{\substack{|m_i|\leqslant h \\ (\mathbf{a},\mathbf{m})\equiv 0(\bmod n)}}{}' \frac{1}{\|\mathbf{m}\|} \\
\leqslant &\sum_{l=1}^{h^s} \frac{(T_{l+1,M}-T_{l,M})}{lM} \leqslant M^{-1}\sum_{l=1}^{h^s} T_{l+1,M}\left(\frac{1}{l}-\frac{1}{l+1}\right)+\frac{T_{h^s+1,M}}{(h^s+1)M} \\
\leqslant &c(s)M^{-1}\sum_{l=1}^{h^s}\frac{(\ln 3lM)^{s-1}}{l}+c(s)M^{-1}(\ln hM)^{s-1} \\
\leqslant &c(s)M^{-1}(\ln hM)^{s}.
\end{aligned}$$

Take $r=1$, $\eta = \dfrac{1}{7M}$ and $h = 7([M]+1)^2$ in Theorem 3.1. Then we have the theorem.

## 3.6 The partial summation formula

**Lemma3.7** *Let $g(\mathbf{m})$ be a non-negative function of $\mathbf{m}$. Then*

$$\begin{aligned}
\sum_{|m_i|\leqslant h}{}' \frac{g(\mathbf{m})}{\|\mathbf{m}\|} \leqslant &\sum_{l=0}^{s}\sum_{\mathbf{i}}\frac{1}{h^l}\sum_{m_{i_{l+1}}=1}^{h}\cdots\sum_{m_{i_s}=1}^{1}\frac{1}{m_{i_{l+1}}^2\cdots m_{i_s}^2} \\
&\times\sum_{|k_{i_1}|\leqslant h}\cdots\sum_{|k_{i_l}|\leqslant h}\ \sum_{|k_{i_{l+1}}|\leqslant m_{i_{l+1}}}\cdots\sum_{|k_{k_{i_s}}|\leqslant m_{i_s}}{}' g(\mathbf{k}),
\end{aligned}$$

*where $\sum_{\mathbf{i}}$ denotes a sum in which $\mathbf{i}=(i_1,\cdots,i_s)$ runs over all the permutations of $(1,\cdots,s)$.*

*Proof.* Since

$$\begin{aligned}
\sum_{|m|\leqslant h}\frac{g(m)}{\bar{m}} &= g(0)+\sum_{m=1}^{h}\frac{1}{m}(g(m)+g(-m)) \\
&= g(0)+\sum_{m=1}^{h}\left(\frac{1}{m}-\frac{1}{m+1}\right)\sum_{1\leqslant k\leqslant m}(g(k)+g(-k))+\frac{1}{h+1}\sum_{|k|\leqslant h}{}' g(k) \\
&\leqslant \sum_{m=1}^{h}\frac{1}{m^2}\sum_{|k|\leqslant m} g(k)+\frac{1}{h}\sum_{|k|\leqslant h} g(k),
\end{aligned}$$

hence

$$\begin{aligned}\sum_{|m_i|\leqslant h}{}' \frac{g(\mathbf{m})}{\|\mathbf{m}\|} \leqslant & \sum_{\substack{|m_i|\leqslant h\\ 1\leqslant i\leqslant s-1}}{}' \frac{1}{\bar{m}_1\cdots\bar{m}_{s-1}}\Bigg(\sum_{m_s=1}^{h}\frac{1}{m_s^2}\sum_{|k_s|\leqslant m_s} g(m_1,\cdots,m_{s-1},k_s)\\ &+\frac{1}{h}\sum_{|k_s|\leqslant h} g(m_1,\cdots,m_{s-1},k_s)\Bigg)\\ &+\sum_{\substack{|m_i|\leqslant h\\ 1\leqslant i\leqslant s-1}} \frac{1}{\bar{m}_1\cdots\bar{m}_{s-1}}\Bigg(\sum_{m_s=1}^{h}\frac{1}{m_s^2}\sum_{|k_s|\leqslant m_s}{}' g(m_1,\cdots,m_{s-1},k_s)\\ &+\frac{1}{h}\sum_{|k_s|\leqslant h}{}' g(m_1,\cdots,m_{s-1},k_s)\Bigg)\leqslant\cdots\\ \leqslant & \sum_{l=0}^{s}\sum_{\mathrm{i}}\frac{1}{h^l}\sum_{m_{i_{l+1}}=1}^{h}\cdots\sum_{m_{i_s}=1}^{h}\frac{1}{m_{i_{l+1}}^2\cdots m_{i_s}^2}\\ &\cdot\sum_{|k_{i_1}|\leqslant h}\cdots\sum_{|k_{i_l}|\leqslant h}\sum_{|k_{i_{l+1}}|\leqslant m_{i_{l+1}}}\cdots\sum_{|k_{i_s}|\leqslant m_{i_s}}{}' g(\mathrm{k}).\end{aligned}$$

The lemma is proved.

## 3.7 The comparison of discrepancies

**Lemma 3.8** *Suppose that the sets* $P_s(k) = (x_1^{(n)}(k), \cdots, x_s^{(n)}(k))(1 \leqslant k \leqslant n)$ *and* $Q_n(k) = (y_1^{(n)}(k), \cdots, y_s^{(n)}(k))(1 \leqslant k \leqslant n)$ *have discrepancies* $D(n)$ *and* $E(n)$ *respectively and that*

$$|x_i^{(n)}(k) - y_i^{(n)}(k)| \leqslant \delta, \quad 1 \leqslant k \leqslant n, 1 \leqslant i \leqslant s. \tag{3.15}$$

*Then*

$$|D(n) - E(n)| \leqslant s\delta.$$

*Proof.* For any $\boldsymbol{\gamma} \in G_s$, set $\boldsymbol{\gamma}' = (\gamma_1', \cdots, \gamma_s')$ and $\boldsymbol{\gamma}'' = (\gamma_1'', \cdots, \gamma_s'')$ where

$$\gamma_i' = \begin{cases} \gamma_i - \delta, & \text{if} \quad \gamma_i - \delta \geqslant 0,\\ 0, & \text{if} \quad \gamma_i - \delta < 0.\end{cases}$$

and

$$\gamma_i'' = \begin{cases} \gamma_i + \delta, & \text{if} \quad \gamma_i + \delta \leqslant 1,\\ 1, & \text{if} \quad \gamma_i + \delta > 1.\end{cases}$$

Let $N_n(\boldsymbol{\gamma})$ and $M_n(\boldsymbol{\gamma})$ denote the numbers of $P_n(k)(1 \leqslant k \leqslant n)$ and $Q_n(k)(1 \leqslant k \leqslant n)$ falling into the region

$$0 \leqslant x_i < \gamma_i, \quad 1 \leqslant i \leqslant s$$

respectively. Then by (3.15),

$$M_n(\gamma') \leqslant N_n(\gamma) \leqslant M_n(\gamma''). \tag{3.16}$$

Let

$$\sigma_1 = \left|\frac{M_n(\gamma')}{n} - |\gamma|\right| \quad \text{and} \quad \sigma_2 = \left|\frac{M_n(\gamma'')}{n} - |\gamma|\right|.$$

Then by (3.16),

$$\left|\frac{N_n(\gamma)}{n} - |\gamma|\right| \leqslant \max(\sigma_1, \sigma_2). \tag{3.17}$$

Since

$$0 \leqslant |\gamma''| - |\gamma| \leqslant \delta + \gamma_1 \left(\prod_{i=2}^{s} \gamma_i'' - \prod_{i=2}^{s} \gamma_i\right) \leqslant \cdots \leqslant s\delta,$$

therefore

$$\sigma_2 \leqslant \left|\frac{M_n(\gamma'')}{n} - |\gamma''|\right| + |\gamma''| - |\gamma| \leqslant E(n) + s\delta.$$

$\sigma_1$ satisfies the same inequality. Hence by (3.17),

$$D(n) \leqslant E(n) + s\delta.$$

Similarly,

$$E(n) \leqslant D(n) + s\delta.$$

The lemma is proved.

## 3.8 Rational approximation and the solutions of the congruence

We use $\mathbf{h} = (h_0, h_1, \cdots, h_s), \mathbf{m} = (m_1, \cdots, m_s)$ and $\mathbf{m}^{(0)} = (m_0, m_1, \cdots, m_s)$ to denote the vectors with integral components, where $h_0 = 1$.

**Lemma3.9** *Suppose that*

$$\langle(\mathbf{r}, \mathbf{m})\rangle \geqslant b\|\mathbf{m}\|^{-a} \tag{3.18}$$

*holds for any* $\mathbf{m} \neq \mathbf{0}$, *where* $a, b$ *are constants satisfying* $s \geqslant a \geqslant 1$, $b > 0$, *and that*

$$\left|\frac{h_i}{n} - \gamma_i\right| \leqslant dn^{-1-g}, \quad 1 \leqslant i \leqslant s, \tag{3.19}$$

*where* $d, g$ *are constants satisfying* $d > 0$, $0 \leqslant g \leqslant 1/s$. *Then there exists a positive constant* $c(b, d, s)(< 1)$ *such that the congruence*

$$(\mathbf{h}, \mathbf{m}^{(0)}) = \sum_{i=0}^{s} h_i m_i \equiv 0(\mathrm{mod} n) \tag{3.20}$$

*has no solution in the domain*

$$\|\mathbf{m}^{(0)}\| \leqslant c(b,d,s)n^{\frac{1+g}{1+a}}, \quad \mathbf{m}^{(0)} \neq \mathbf{0}. \tag{3.21}$$

*Proof* Suppose that $\mathbf{m}^{(0)}(\neq \mathbf{0})$ is a solution of (3.20). If $\mathbf{m} = \mathbf{0}$, then $m_0 \neq 0$. By (3.20), we have $m_0 \equiv 0(\mathrm{mod} n)$. Hence $\|\mathbf{m}^{(0)}\| \geqslant n$ and so $\mathbf{m}^{(0)}$ not belongs to the domain (3.21). Consequently, we may suppose that $\mathbf{m} \neq \mathbf{0}$. If

$$\|\mathbf{m}\| \geqslant \left(\frac{b}{2ds}\right)^{\frac{1}{1+a}} n^{\frac{1+g}{1+a}},$$

then

$$\|\mathbf{m}^{(0)}\| \geqslant \left(\frac{b}{2ds}\right)^{\frac{1}{1+a}} n^{\frac{1+g}{1+a}}$$

and so we have the theorem. Now, suppose that

$$\|\mathbf{m}\| < \left(\frac{b}{2ds}\right)^{\frac{1}{1+a}} n^{\frac{1+g}{1+a}}.$$

Since

$$\langle \alpha - \beta \rangle \geqslant \langle \alpha \rangle - \langle \beta \rangle,$$

we have

$$\begin{aligned}\frac{|m_0|}{n} \geqslant \left\langle \frac{m_0}{n} \right\rangle = \left\langle \frac{1}{n}\sum_{i=1}^{s} h_i m_i \right\rangle \geqslant \langle (\boldsymbol{\gamma}, \mathbf{m}) \rangle \\ - \left\langle \sum_{i=1}^{s} \left(\frac{h_i}{n} - \gamma_i\right) m_i \right\rangle \geqslant \frac{b}{\|\mathbf{m}\|^a} - \frac{ds}{n^{1+g}}\|\mathbf{m}\|\end{aligned}$$

by (3.18) and (3.19). Hence

$$\begin{aligned}\|\mathbf{m}^{(0)}\| \geqslant & \frac{nb}{\|\mathbf{m}\|^{a-1}} - \frac{ds}{n^g}\|\mathbf{m}\|^2 \geqslant b^{\frac{2}{1+a}}(2ds)^{\frac{-1+a}{1+a}} n^{1+\frac{(1-a)(1+g)}{1+a}} \\ & - \left(\frac{b}{2}\right)^{\frac{2}{1+a}} (ds)^{\frac{-1+a}{1+a}} n^{\frac{2(1+g)}{1+a}-g} \geqslant \frac{1}{2}(2ds)^{\frac{-1+a}{1+a}} b^{\frac{2}{1+a}} n^{\frac{1+g}{1+a}}.\end{aligned}$$

The lemma is proved.

## 3.9 The rational approximation and the discrepancy

In this section, we shall prove the following three theorems.

**Theorem3.3** *Suppose that* (3.18) *holds for* $\mathbf{m} \neq \mathbf{0}$, *where* $a, b$ *are constants satisfying* $1 \leqslant a \leqslant 1 + \dfrac{1}{2s}$ *and* $b > 0$. *Then the set*

$$P_n(k) = (\{\gamma_1 k\}, \cdots, \{\gamma_s k\}), \quad 1 \leqslant k \leqslant n$$

*has discrepancy*

$$D(n) \leqslant c(b, s)n^{-1+2s(a-1)}(\ln n)^{1+s\delta_{1,a}}.$$

**Theorem 3.4** *Suppose that* (3.19) *holds, where* $d, g$ *are constants satisfying* $d > 0$ *and* $0 \leqslant g \leqslant \frac{1}{s}$. *Then under the assumption of Theorem* 3.3, *the set*

$$\left(\left\{\frac{k}{n}\right\}, \left\{\frac{h_1 k}{n}\right\}, \cdots, \left\{\frac{h_s k}{n}\right\}\right), \quad 1 \leqslant k \leqslant n$$

*has discrepancy*

$$D(n) \leqslant c(b, d, s)n^{-\frac{-1+g}{1+a}}(\ln n)^{s+1}.$$

**Theorem3.5** *Suppose that* $q$ *is an integer satisfying* $1 \leqslant q \leqslant n^{\frac{1+g}{n^{2-2s(a-1)}}}$. *Then under the assumption of Theorem* 3.4, *the set*

$$\left(\left\{\frac{h_1 k}{n}\right\}, \cdots, \left\{\frac{h_s k}{n}\right\}\right), \quad 1 \leqslant k \leqslant q \tag{3.22}$$

*has discrepancy*

$$D(q) \leqslant c(b, d, s)q^{-1+2s(a-1)}(\ln 3q)^{1+s\theta_{1,a}}.$$

To prove these theorems, we shall need

**Lemma3.10** *Let* $\delta$ *be a real number. Then*

$$\left|\sum_{k=1}^{n} e^{2\pi i \delta k}\right| \leqslant \min\left(n, \frac{1}{2\langle\delta\rangle}\right)$$

(Cf. Hua Loo Keng [2], Chap. 7).

**Lemma 3.11**

$$\sum_{m=1}^{n} \frac{1}{m^\alpha} \leqslant \begin{cases} \dfrac{n^{1-\alpha}}{1-2}, & if \quad 0 \leqslant \alpha < 1, \\ \dfrac{(n+1)^{1-\alpha}}{1-\alpha}, & if \quad \alpha < 0. \end{cases}$$

*Proof.* For $0 \leqslant \alpha < 1$,

$$\sum_{m=1}^{n} \frac{1}{m^\alpha} \leqslant 1 + \int_1^n \frac{dt}{t^\alpha} = 1 + \frac{n^{1-\alpha}}{1-\alpha} - \frac{1}{1-\alpha} \leqslant \frac{n^{1-\alpha}}{1-\alpha}.$$

The other inequality may be proved similarly.

**Lemma3.12** *Let* $Q = [2^{sa}\|\mathbf{m}\|^a b^{-1}] + 1$, *where* $\mathbf{m} \neq \mathbf{0}$. *Then under the assumption of Theorem* 3.3, *there is at most* 1 *point* $(\mathbf{k}, \gamma) = \sum_{i=1}^{s} k_i \gamma_i$ *in any interval of the type* $(P, P + Q^{-1}]$, *where* $k$ *is a vector with integral components which satisfy* $|k_i| \leqslant |m_i| (1 \leqslant i \leqslant s)$.

*Proof.* If there are two points $(\mathbf{k}'\boldsymbol{\gamma})$ and $(\mathbf{k}'',\boldsymbol{\gamma})$ in the interval $(P, P+Q^{-1}]$, where $\mathbf{k}' \neq \mathbf{k}''$, $|k_i'| \leqslant |m_i|$ and $|k_i''| \leqslant |m_i| (1 \leqslant i \leqslant s)$, then

$$\langle(\mathbf{k}'-\mathbf{k}'',\boldsymbol{\gamma})\rangle \leqslant Q^{-1}$$

On the other hand from (3.18),

$$\langle(\mathbf{k}'-\mathbf{k}'',\boldsymbol{\gamma})\rangle > b\|\mathbf{k}'-\mathbf{k}''\|^{-a} \geqslant 2^{-sa}b\|\mathbf{m}\|^{-a} > Q^{-1}$$

which gives a contradiction. Then lemma is proved.

**Lemma 3.13** *Under the assumption of Lemma* 3.12,

$$\sum_{|k_i|\leqslant|m_i|}{}' \frac{1}{\langle(\mathbf{k},\boldsymbol{\gamma})\rangle} \leqslant 4Q\ln 3Q.$$

*Proof.* Divide the interval (0,1] into $Q$ subintervals

$$I_l = \left(\frac{l}{Q}, \frac{l+1}{Q}\right], \quad l = 0, 1, \cdots, Q-1.$$

None of the points $(\mathbf{k},\boldsymbol{\gamma})$ is contained in $I_0$, where $\mathbf{k} \neq \mathbf{0}$ and $|k_i| \leqslant |m_i| (1 \leqslant i \leqslant s)$. Otherwise be (3.18),

$$Q^{-1} \geqslant \langle(\mathbf{k},\boldsymbol{\gamma})\rangle \geqslant b\|\mathbf{k}\|^{-a} \geqslant b\|\mathbf{m}\|^{-a} > Q^{-1}$$

which gives a contradiction. It follows by Lemma 3.12 that there is at most 1 point $(\mathbf{k},\boldsymbol{\gamma})$ in any interval $I_l$, where $l \geqslant 1$. Hence

$$\sum_{|k_i|\leqslant|m_i|}{}' \frac{1}{\langle(\mathbf{k},\mathbf{r})\rangle} < 4\sum_{k=1}^{Q-1}\frac{Q}{k} < 4Q(1+\text{In}Q) < 4Q\text{In}3Q.$$

The lemma follows.

**Lemma 3.14** *Let $h$ be an integer $\geqslant 2$. Then under the assumption of Theorem*3.3,

$$\sum_{|m_i|\leqslant h}{}' \frac{1}{\|\mathbf{m}\|\langle(\mathbf{m},\boldsymbol{\gamma})\rangle} \leqslant c(b,s)h^{s(a-1)}(\ln h)^{1+s\delta_{1,a}}.$$

*Proof.* By Lemmas 3.2, 3.7, 3.11 and 3.13,

$$\sum_{|m_i|\leqslant h}{}' \frac{1}{\|\mathbf{m}\|\langle(\mathbf{m},\boldsymbol{\gamma})\rangle} \leqslant \sum_{l=0}^{s}\sum_{i}\frac{1}{h^l}\sum_{m_{i_{l+1}}=1}^{h}\cdots\sum_{m_{i_s}=1}^{h}\frac{1}{m_{i_{l+1}}^2\cdots m_{i_s}^2}$$
$$\times \sum_{|k_{i_1}|\leqslant h}\cdots\sum_{|k_{i_l}|\leqslant h}\sum_{|k_{i_{l+1}}|\leqslant m_{i_{l+1}}}\cdots\sum_{|k_{i_s}|\leqslant m_{i_s}}{}' \frac{1}{\langle(\mathbf{k},\boldsymbol{\gamma})\rangle}$$

$$\leqslant \sum_{l=0}^{s} \sum_{i} \frac{1}{h^l} \sum_{m_{i_{l+1}}=1}^{h} \cdots \sum_{m_{i_s}=1}^{h} \frac{1}{m_{i_{l+1}}^2 \cdots m_{i_s}^2} c(b,s) h^{la} (m_{i_{l+1}} \cdots m_{i_s})^a \ln h$$
$$\leqslant c(b,s) h^{s(a-1)} (\ln h)^{1+s\delta_{1,a}}.$$

The lemma is proved.

The proof of Theorem3.3. By Lemmas 3.10 and 3.14,

$$\sum_{|m_i| \leqslant h}{}' \frac{1}{\|\mathbf{m}\|} \left| \frac{1}{n} \sum_{k=1}^{n} e^{2\pi i (\mathbf{m}, \boldsymbol{\gamma}) k} \right| \leqslant \frac{1}{2n} \sum_{|m_i| \leqslant h}{}' \frac{1}{\|\mathbf{m}\| \langle (\mathbf{m}, \boldsymbol{\gamma}) \rangle}$$
$$\leqslant c(b,s) n^{-1} h^{s(a-1)} (\ln h)^{1+s\delta_{1,a}}.$$

Take $r = 1$, $\eta = \dfrac{1}{7n}$ and $h = 8n^2$ in Theorem 3.1. Then we have the theorem.

The proof of Theorem 3.5. From Theorem 3.3, the set

$$(\{\gamma_1 k\}, \cdots, \{\gamma_s k\}), \quad 1 \leqslant k \leqslant q$$

has discrepancy $\leqslant c(b,s) q^{-1+2s(a-1)} (\ln 3q)^{1+s\delta_{1,a}}$. Hence by (3.19) and Lemma 3.8. The discrepancy $D(q)$ of the set (3.22) satisfies

$$D(q) \leqslant c(b,s) q^{-1+2s(a-1)} (\ln 3q)^{1+s\delta_{1,a}} + sdqn^{-1-g}$$
$$\leqslant c(b,d,s) q^{-1+2s(a-1)} (\ln 3q)^{1+s\delta_{1,a}}.$$

The theorem is proved.

Theorem 3.4 may be derived from Theorem 3.2 and Lemma 3.9 immediately.

## 3.10 The lower estimate of discrepancy

**Theorem3.6** *Suppose that $s \geqslant 2$. Then for any set of points $P_n(k) (1 \leqslant k \leqslant n)$,*

$$D(n) > 2^{-2s-4} (s-1)^{-\frac{s-1}{2}} n^{-1} (\log_2 n)^{\frac{s-1}{2}}.$$

Evidently, Theorem3.6 is a consequence of the following:

**Theorem3.7** *Suppose that $s \geqslant 2$. Then for any set of points $P_n(k) (1 \leqslant k \leqslant n)$,*

$$\int_{G_s} (N_n(\mathbf{x}) - n|\mathbf{x}|)^2 d\mathbf{x} > 2^{-4s-8} (s-1)^{-(s-1)} (\log_2 n)^{s-1}.$$

We may suppose that $s = 2$, since the proof of the case $s > 2$ is completely similar. Denote

$$P_n(k) = (X_k, Y_k), \quad 1 \leqslant k \leqslant n.$$

Any point $x$ in the interval $0 \leqslant x \leqslant 1$ can be represented uniquely by the series

$$x = \frac{x_1}{2} + \frac{x_2}{2^2} + \cdots, \quad x_i = 0 \quad \text{or} \quad 1.$$

Let

$$\tau_r(x) = (-1)^{x_r}, \quad r = 1, 2, \cdots.$$

Let $m$ be an integer $> 2$ which will be determined later. For any $(x, y) \in G_2$ we define $F_r(x, y)(r = 1, \cdots, m-1)$ as follows. If there is a $k(1 \leqslant k \leqslant n)$ such that

$$x_1(X_k) = x_1(x), \cdots, x_{r-1}(X_k) = x_{r-1}(x),$$

$$y_1(Y_k) = y_1(y), \cdots, y_{m-r-1}(Y_k) = y_{m-r-1}(y), \tag{3.23}$$

then $F_r(x, y) = 0$, otherwise

$$F_r(x, y) = \tau_r(x)\tau_{m-r}(y).$$

**Lemma3.15** *Suppose that $u$ is a non-negative integer. Then*

$$\int_{u2^{-r+1}}^{(u+1)2^{-r+1}} F_r(x, y)dx = 0 \tag{3.24}$$

*and*

$$\int_{u2^{-m+r+1}}^{(u+1)2^{-m+r+1}} F_r(x, y)dy = 0. \tag{3.25}$$

*Proof.* For any given $y$ and $x \in [u2^{-r+1}, (u+1)2^{-r+1})$, the $x_1, \cdots, x_{r-1}$ in the binary representation of $x$ are the same, but

$$x_r = \begin{cases} 0, & \text{if} \quad x \in [u2^{-r+1}, u2^{-r+1} + 2^{-r}), \\ 1, & \text{if} \quad x \in [u2^{-r+1} + 2^{-r}, (u+1)2^{-r+1}). \end{cases}$$

hence we have (3.24). Then proof of (3.25) is similar.

**Lemma 3.16** *Let*

$$F(x, y) = \sum_{r=1}^{m-1} F_r(x, y). \tag{3.26}$$

*Then*

$$\int_0^1 \int_0^1 xyF(x, y)dxdy \geqslant (m-1)2^{-2m}(2^{m-2} - n).$$

*Proof.* It is sufficient to prove that

$$\int_0^1 \int_0^1 xyF_r(x, y)dxdy \geqslant 2^{-2m}(2^{m-2} - n), \quad r = 1, 2, \cdots, m-1.$$

Divide the intervals of the integrals of $x$ and $y$ into $2^{r-1}$ and $2^{m-r-1}$ equal parts respectively. Then the above integral is equal to

$$\Sigma^* \int_{-2^{-r}}^{2^{-r}} \int_{-2^{-m+r}}^{2^{-m+r}} (\xi + x)(\eta + y)(\text{sign}x)(\text{sign}y)dxdy, \tag{3.27}$$

where

$$\xi = \frac{x_1}{2} + \cdots + \frac{x_{r-1}}{2^{r-1}} + \frac{1}{2^r},$$

$$\eta = \frac{y_1}{2} + \cdots + \frac{y_{m-r-1}}{2^{m-r-1}} + \frac{1}{2^{m-r}}$$

and $\Sigma^*$ denotes a sum of $x_1, \cdots, x_{r-1}, y_1, \cdots, y_{m-r-1}$ such that (3.23) is not satisfied for any $k (1 \leqslant k \leqslant n)$.

Since

$$\int_{-a}^{a} (\xi + x)\text{sign}x dx = \int_{-a}^{a} x\text{sign}x dx = a^2,$$

(3.27) is equal to

$$\Sigma^* 2^{-2m}.$$

Since the total number of the sets of $x_1, \cdots, x_{r-1}, y_1, \cdots, y_{m-r-1}$ is $2^{m-2}$ in which the number of sets such that (3.23) are satisfied is at most $n$, hence the total number of terms of $\Sigma^*$ is no less than $2^{m-2} - n$. The lemma follows.

**Lemma 3.17**

$$\int_0^1 \int_0^1 F(x, y)^2 dx dy \leqslant m - 1.$$

*Proof.* By (3.26),

$$\int_0^1 \int_0^1 F(x, y)^2 dx dy = \sum_{r=1}^{m-1} \int_0^1 \int_0^1 F_r(x, y)^2 dx dy + 2 \sum_{1 \leqslant r < s \leqslant m-1} \int_0^1 \int_0^1 F_r(x, y) F_s(x, y) dx dy.$$

Since $|F_r(x, y)| \leqslant 1$, the fires term of the right hand side does not exceed $m - 1$. Now we proceed to prove that the second term is equal to zero. For any given $y$, divide the interval of the integral of $x$ into $2^{s-1}$ equal parts. Similar to (3.24), the integrals over these subintervals are all equal to zero. Hence we have the lemma.

**Lemma3.18**

$$\int_{X_k}^1 \int_{Y_k}^1 F(x, y) dx dy = 0, \quad 1 \leqslant k \leqslant n.$$

*Proof.* It is sufficient to prove that

$$\int_{X_k}^1 \int_{Y_k}^1 F_r(x, y) dx dy = 0, \quad r = 1, \cdots, m - 1.$$

Divide the integral into four parts

$$\int_{X_k}^X \int_{Y_k}^Y + \int_X^1 \int_{Y_k}^Y + \int_{X_k}^X \int_Y^1 + \int_X^1 \int_Y^1, \tag{3.28}$$

where $X$ is the least integer $\geqslant X_k$ and also a multiple of $2^{-r+1}$ and where $Y$ is the least integer $\geqslant Y_k$ and also a multiple of $2^{-m+r+1}$. Since (3.23) holds in the rectangle

$$X_k \leqslant x < X, \quad Y_k \leqslant y < Y,$$

the first integral of (3.28) is equal to zero. It follows by Lemma 3.15 that the other integrals are all equal to zero too. The lemma is proved.

The proof of Theorem 3.7. From Lemma 3.18,

$$\begin{aligned}\int_0^1\int_0^1 N_n(x,y)F(x,y)dxdy &= \int_0^1\int_0^1 F(x,y)\left(\sum_{\substack{k=1\\ X_k<x,Y_k<y}}^{n} 1\right)dxdy\\ &= \sum_{k=1}^{n}\int_{X_k}^1\int_{Y_k}^1 F(x,y)dxdy = 0.\end{aligned}$$

Hence by Lemma 3.16,

$$\begin{aligned}\int_0^1\int_0^1 (nxy - N_n(x,y))F(x,y)dxdy &= n\int_0^1\int_0^1 xyF(x,y)dxdy\\ &\geqslant n(m-1)2^{-2m}(2^{m-2}-n).\end{aligned}$$

Let $m$ be the integer such that

$$8n < 2^m \leqslant 16n.$$

Then

$$\begin{aligned}\int_0^1\int_0^1 (nxy-N_n(x,y))^2dxdy &\geqslant \left(\int_0^1\int_0^1 (nxy-N_n(x,y))F(x,y)dxdy\right)^2\\ &\left(\int_0^1\int_0^1 F(x,y)^2dxdy\right)^{-1} \geqslant (m-1)(n2^{-2m}(2^{m-2}-n))^2\end{aligned}$$

by Schwarz inequality and Lemma 3.17. Since

$$2^{m-2}-n > n, \quad n^2 2^{-2m} \geqslant 2^{-s}, \quad m > \log_2 n + 3,$$

hence

$$\int_0^1\int_0^1 (N_n(x,y) - nxy)^2dxdy > 2^{-16}\log_2 n.$$

The theorem is proved.

If the sequence of sets $P_{nl}(k)(n_1 < n_2 < \cdots)$ have discrepancy

$$D(n) = O(n^{-1+\varepsilon}),$$

where the constant implied by the symbol "$O$" depends only on $\varepsilon$, the sequence $P_{n_l}(k)(n_1 < n_2 < \cdots)$ (or simply the set $P_n(k)$) is said to be best uniformly distributed.

**Notes**

The definition of uniform distribution was first given by H. Weyl [1].

Lemma 3.1: Cf. I. M. Vinogradov [1].

Theorem 3.1 in a slightly different form was proved by P. Erdös, and P. Turán [1] for $s = 1$ and J. F. Koksma [2] for $s > 1$ (Cf. also E. Hlawka [2] and Hua Loo Keng and Wang Yuan [7]).

Lemmas 3.3, 3.4 and 3.5: Cf. N. S. Bahvalov [1] and also Hua Loo Keng and Wang Yuan [7].

Theorem 3.3 was proved independently by Hua Loo Keng and Wang Yuan [7] and H. Niederreiter [1] in a slightly different form.

Concerning the lower estimation of the discrepancy, T. Van Aardenne-Ehrenfest [1] first proved that $D(n) > \frac{c\ln\ln n}{n\ln\ln\ln n}$. Theorem 3.6 is due to K. F. Roth [1] whose result was improved to $D(n) > \frac{c\ln n}{n}$ for $s = 2$ by W. M. Schmidt [3].

# Chapter 4

# Estimation of Discrepancy

## 4.1 The set of equi-distribution

Suppose that $s \geqslant 2, m_1, \cdots, m_s$ are $s$ positive integers, $n = m_1 \cdots, m_s$ and $m = \min(m_1, \cdots, m_s)$. The set

$$\left(\frac{l_1}{m_1}, \cdots, \frac{l_s}{m_s}\right), \quad 0 \leqslant l_i < m_i, 1 \leqslant i \leqslant s \tag{4.1}$$

is called the set of equi-distribution.

**Theorem 4.1** *The set* (4.1) *has discrepancy*

$$D(n) \leqslant 2^s m^{-1}.$$

*Proof.* Let $\gamma \in G_s$. Since the number of $l_i$ satisfying

$$\frac{l_i}{m} < \gamma_i, \quad l_i = 0, 1, \cdots$$

is equal to $[m_i\gamma_i]$ or $[m_i\gamma_i] + 1$, therefore

$$N_n(\boldsymbol{\gamma}) = ([m_1\gamma_1] + \theta_1) \cdots ([m_s\gamma_s] + \theta_s),$$

where $\theta_i = 0$ or $1$ $(1 \leqslant i \leqslant s)$. Hence

$$\begin{aligned} 0 \leqslant \frac{N_n(\boldsymbol{\gamma})}{n} - |\gamma| &\leqslant \prod_{i=1}^{s} \frac{(m_i\gamma_i + 1)}{m_i} - \prod_{i=1}^{s} \gamma_i \\ &= \prod_{i=1}^{s} \left(\gamma_i + \frac{1}{m_i}\right) - \prod_{i=1}^{s} \gamma_i \leqslant 2^s m^{-1}. \end{aligned}$$

The theorem is proved.

**Theorem 4.2** *The discrepancy of the set* (4.1) *satisfies*

$$D(n) \geqslant \frac{1}{2m} \geqslant 2^{-1} n^{-1/s}.$$

*Proof.* Without loss of generality, we may suppose that $m = m_1$. Set $\gamma = \left(\frac{1}{2m}, 1, \cdots, 1\right)$. Then

$$N_n(\gamma) = m_2 \cdots m_s = \frac{n}{m}.$$

Hence

$$\left|\frac{N_n(\gamma)}{n} - |\gamma|\right| = \frac{1}{m} - \frac{1}{2m} = \frac{1}{2m} \geqslant 2^{-1}n^{-1/s}.$$

The theorem is proved.

Clearly, the best case for the equi-distribution is to choose $m_1 = \cdots = m_s = m$, i.e.,

$$\left(\frac{l_1}{m}, \cdots, \frac{l_s}{m}\right), \quad 0 \leqslant l_1, \cdots, l_s < m.$$

And it follows from Theorem 4.2 that the discrepancy $D(n)$ of the set of equi-distribution increases rapidly when $s$ increases.

## 4.2 The Halton theorem

**Lemma 4.1** *The number of solutions of the congruence*

$$x \equiv a (\text{mod } m), \quad 1 \leqslant x \leqslant n$$

*is equal to* $\left[\frac{n}{m}\right]$ *or* $\left[\frac{n}{m}\right] + 1$.

*Proof.* Since the congruence has exactly 1 solution in any $m$ consecutive integers, the lemma follows.

**Lemma 4.2** *Suppose that* $m_1, \cdots, m_s$ *are positive integers which are relatively prime to each other. Then the system of congruences*

$$x \equiv a_i (\text{mod } m_i), \quad 1 \leqslant i \leqslant s$$

*has a unique solution* mod $m_1 \cdots m_s$ (Cf. Hua Loo Keng [2], Chap. 2).

Let $r$ be an integer$>1$. Then any positive integer $k$ can be represented uniquely as

$$k = k_0 + k_1 r + \cdots + k_M r^M, \quad 0 \leqslant k_i \leqslant r - 1.$$

Which is denoted in digits by

$$k = k_M \cdots k_1 k_0$$

and any number $h$ in the interval [0,1] can be represented uniquely by

$$h = \frac{h_0}{r} + \frac{h_1}{r^2} + \cdots + \frac{h_M}{r^{M+1}} + \cdots, \quad 0 \leqslant h_i \leqslant r - 1$$

which is denoted by

$$h = 0.h_0h_1 \cdots h_M \cdots .$$

For any given positive integer $k = k_0 + k_1r + \cdots + k_Mr^M$, where it corresponds to a number

$$\varphi_r(k) = \frac{k_0}{r} + \frac{k_1}{r^2} + \cdots + \frac{k_M}{r^{M+1}}.$$

If $k_M \neq 0$, then

$$r^M \leqslant k < r^{M+1}$$

and so

$$M = \left[\frac{\ln k}{\ln r},\right]$$

where $M + 1$ is called the number of digits of $k$.

**Lemma 4.3** *Suppose that $n > r$. Then the set*

$$\varphi_r(k), \quad k = 1, 2, \cdots, n$$

*has discrepancy*

$$D(n) < \left(\frac{r \ln rn}{\ln r}\right)n^{-1}.$$

*Proof.* Let $\alpha$ satisfy $0 < \alpha \leqslant 1$. Then $\alpha$ may be written as

$$\alpha = 0.a_0a_1 \cdots a_M \cdots .$$

Without loss of generality, we may suppose that $\alpha$ does not terminate. Otherwise if $\alpha = 0.a_0a_1 \cdots a_M$, where $a_M \neq 0$, then $\alpha$ may be written as $\alpha = 0.a_0a_1 \cdots a_{M-1}a'_M a'_{M+1} \cdots$, where $a'_M = a_M - 1, a'_i = r - 1 (i = M + 1, M + 2, \cdots)$. If $\varphi_r(k) < \alpha$, then the integer $k$ satisfies one of the following conditions:

1) $k_0 < a_0$,

2) $k_0 = a_0, k_1 < a_1$,

3) $k_0 = a_0, k_1 = a_1, k_2 < a_2$,

$\cdots$

Since the number of digits of $k$ does not exceed $\left[\frac{\ln n}{\ln \mathrm{r}}\right] + 1$ for $1 \leqslant k \leqslant n$, we may write $M = \left[\frac{\ln n}{\ln r}\right]$. The final two steps are

$M + 1$) $k_0 = a_0, k_1 = a_1, \cdots, k_{M-1} = a_{M-1}, k_M < a_M$,

$M + 2$) $k_0 = a_0, k_1 = a_1, \cdots, k_{M-1} = a_{M-1}, k_M = a_M$.

Write these formulas in the forms of congruences

1) $k \equiv k_0 (\text{mod } r), 0 \leqslant k_0 < a_0,$

2) $k \equiv a_0 + k_1 r (\text{mod } r^2), 0 \leqslant k_1 < a_1,$

3) $k \equiv a_0 + a_1 r + k_2 r^2 (\text{mod } r^3), 0 \leqslant k_2 < a_2,$

...

$M+1$) $k \equiv a_0 + a_1 r + \cdots + a_{M-1} r^{M-1} + k_M r^M (\text{mod } r^{M+1}) \quad 0 \leqslant k_M < a_M,$

$M+2$) $k \equiv a_0 + a_1 r + \cdots + a_{M-1} r^{M-1} + a_M r^M (\text{mod } r^{M+2}).$

By Lemma 4.1, The numbers of solutions of the above congruences for $1 \leqslant k \leqslant n$ are equal to

$$\begin{aligned} &a_0\Big(\Big[\frac{n}{r}\Big] + \theta\Big) \\ &a_1\Big(\Big[\frac{n}{r^2} + \theta\Big]\Big), \\ &\cdots\cdots \\ &a_M\Big(\Big[\frac{n}{r^{M+1}}\Big] + \theta\Big), \\ &\Big(\Big[\frac{n}{r^{M+2}}\Big] + \theta\Big) = \theta \end{aligned}$$

respectively, where $\theta$ satisfies $0 \leqslant \theta \leqslant 1$. (But not always having the same value.)

Let $N_n(\alpha)$ denote the number of $k$ satisfying $\varphi_r(k) < \alpha (1 \leqslant k \leqslant n)$.

Then

$$N_n(\alpha) = a_0\Big(\Big[\frac{n}{r}\Big] + \theta\Big) + a_1\Big(\Big[\frac{n}{r^2}\Big] + \theta\Big) + \cdots a_M\Big(\Big[\frac{n}{r^{M+1}}\Big] + \theta\Big) + \theta$$

and

$$\begin{aligned} &\Big|N_n(\alpha) - a_0\frac{n}{r} - a_1\frac{n}{r^2} - \cdots - a_M\frac{n}{r^{M+1}}\Big| \\ &\leqslant a_0\Big|\frac{n}{r} - \Big[\frac{n}{r}\Big] - \theta\Big| + a_1\Big|\frac{n}{r^2} - \Big[\frac{n}{r^2}\Big] - \theta\Big| + \cdots \\ &\quad + a_M\Big|\frac{n}{r^{M+1}} - \Big[\frac{n}{r^{M+1}}\Big] - \theta\Big| + \theta \\ &\leqslant a_0 + a_1 + \cdots + a_M + 1. \end{aligned}$$

Hence

$$|N_n(\alpha) - \alpha n| \leqslant a_0 + a_1 + \cdots + a_M + 1 + \Big(\frac{a_{M+1}}{r^{M+2}} + \frac{a_{M+2}}{r^{M+3}} + \cdots\Big)n. \tag{4.2}$$

Since

$$\begin{aligned} \frac{a_{M+1}}{r^{M+2}} + \frac{a_{M+2}}{r^{M+3}} + \cdots &\leqslant (r-1)\Big(\frac{1}{r^{M+2}} + \frac{1}{r^{M+3}} + \cdots\Big) \\ &= \frac{r-1}{r^{M+2}}\Big(1 - \frac{1}{r}\Big)^{-1} = \frac{1}{r^{M+1}} < \frac{1}{n}, \end{aligned}$$

we have

$$|N_n(\alpha) - \alpha n| \leqslant (M+1)(r-1)+2$$
$$\leqslant \Big(\frac{\ln n}{\ln r}+1\Big)(r-1)+2 \leqslant \frac{r \ln rn}{\ln r}$$

for $n > r$. The lemma is proved.

**Theorem 4.3** *Suppose that $r_i (1 \leqslant i \leqslant s)$ are $s$ integers$> 1$ which are relatively prime to each other and that $n >\max(r_1, \cdots, r_s)$. Then the set*

$$(\varphi_{r_1}(k), \cdots, \varphi_{r_s}(k)), \quad 1 \leqslant k \leqslant n \tag{4.3}$$

*has discrepancy*

$$D(n) \leqslant \Big(\prod_{i=1}^{s} \frac{r_i \ln r_i n}{\ln r_i}\Big) n^{-1}.$$

*Proof.* We may suppose that $s = 2$, since the proof is similar for the case $s > 2$. Suppose that $r$ ant $t$ are integers$> 1$ and $(r, t) = 1$ and that $0 \leqslant \beta < 1$ and its expansion in the scale of $t$ is

$$\beta = 0.b_0 b_1 \cdots .$$

Let $N_n(\alpha, \beta)$ denote the number of integers $k$ satisfying

$$\varphi_r(k) < \alpha, \quad \varphi_t(k) < \beta, \quad 1 \leqslant k \leqslant n.$$

Let $L = \left[\frac{\ln n}{\ln t}\right]$. Then the integers in the interval $1 \leqslant k \leqslant n$ can be represented uniquely by

$$k = l_0 + l_1 t + \cdots + l_L t^L, \quad 0 \leqslant l_i \leqslant t-1.$$

Similar to the congruences 1),2),$\cdots$, M+2), we have a system of congruences

1)′ $k \equiv l_0 (\text{mod } t), 0 \leqslant l_0 < b_0,$

2)′ $k \equiv b_0 + l_1 t (\text{mod } t^2), 0 \leqslant l_1 < b_1,$

...

$L+1)'$ $k \equiv b_0 + b_1 t + \cdots + b_{L-1} t^{L-1} + l_L t^L (\text{mod } t^{L+1}), 0 \leqslant l_L < b_L,$

$L+2)'$ $k \equiv b_0 + b_1 t + \cdots + b_{L-1} t^{L-1} + b_L t^L (\text{mod } t^{L+2}).$

Since if follows by Lemmas 4.1 and 4.2 that the number of integers in $1 \leqslant k \leqslant n$ which satisfy $m)$ and $l)'$ is

$$a_{m-1} b_{l-1} \Big( \Big[\frac{n}{r^m t^l}\Big] + \theta \Big),$$

so $N_n(\alpha, \beta)$ is equal to

$$\sum_{m=1}^{M+1}\sum_{l=1}^{L+1} a_{m-1}b_{l-1}\Big(\Big[\frac{n}{r^m t^l}\Big]+\theta\Big)+\sum_{m=1}^{M+1} a_{m-1}\theta+\sum_{l=1}^{L+1} b_{l-1}\theta+\theta$$

and so

$$\begin{aligned}&\Big|N_n(\alpha,\beta)-\sum_{m=1}^{M+1}\sum_{l=1}^{L+1} a_{m-1}b_{l-1}\frac{n}{r^m t^l}\Big|\\ &\quad\leqslant \sum_{m=1}^{M+1}\sum_{l=1}^{L+1} a_{m-1}b_{l-1}+\sum_{m=1}^{M+1} a_{m-1}+\sum_{l=1}^{L+1} b_{l-1}+1\\ &\quad=\Big(\sum_{m=1}^{M+1} a_{m-1}+1\Big)\Big(\sum_{l=1}^{L+1} b_{l-1}+1\Big).\end{aligned}$$

Consequently, we have

$$\begin{aligned}|N_n(\alpha,\beta)-\alpha\beta n| &\leqslant \Big(\sum_{m=1}^{M+1} a_{m-1}+1\Big)\Big(\sum_{l=1}^{L+1} b_{l-1}+1\Big)\\ &+n\sum_{m=M+2}^{\infty}\frac{a_{m-1}}{r^m}+n\sum_{l=L+2}^{\infty}\frac{b_{l-1}}{t^l}+n^{-1}\\ &\leqslant((r-1)(M+1)+1)((t-1)(L+1)+1)+3\\ &\leqslant((r-1)(M+1)+2)((t-1)(L+1)+2)\\ &\leqslant\Big(\frac{r\ \ln rn}{\ln r}\Big)\Big(\frac{t\ \ln tn}{\ln t}\Big).\end{aligned}\tag{4.4}$$

The theorem is proved.

**Lemma 4.4** *Suppose that $n > r^2$. Then under the assumption of Lemma 4.3,*

$$\frac{1}{q}\sum_{l=1}^{q}\Big|\frac{1}{n}N_n\Big(\frac{l}{q}\Big)-\frac{l}{q}\Big|\leqslant\Big(\frac{\ln rn}{2\ \ln r}\Big)n^{-1}.$$

*Proof.* Let

$$1-\alpha=0.a_0'a_1'\cdots a_M'\cdots.$$

Then

$$a_v+a_v'=r-1,\quad v=0,1,\cdots$$

and by (4.2),

$$\begin{aligned}
&|N_n(\alpha) - \alpha n| + |N_n(1-\alpha) - (1-\alpha)n| \\
&\quad \leqslant (r-1)(M+1) + 2 + (r-1)n \sum_{l=M+2}^{\infty} \frac{1}{r^l} \\
&\quad \leqslant (r-1)(M+1) + 2 + \frac{n}{r^{M+1}} \\
&\quad \leqslant (r-1)(M+1) + 3 \leqslant \frac{r \ln rn}{\ln r}.
\end{aligned}$$

Hence

$$\frac{1}{q}\sum_{l=1}^{q}\left|\frac{1}{n}N_n\Big(\frac{l}{q}\Big) - \frac{l}{q}\right| \leqslant \Big(\frac{r \ln rn}{2\ln r}\Big)n^{-1}.$$

The lemma is proved.

**Theorem 4.4** *Suppose that* $n > \max\,(r_1^4, \cdots r_s^4)$. *Then under the assumption of Theorem* 4.3,

$$\begin{aligned}
&\frac{1}{q^s}\sum_{l=1}^{q}\cdots\sum_{l_s=1}^{q} q^{\delta_{l_1,q}+\cdots+\delta_{l_s,q}}\left|\frac{1}{n}N_n\Big(\frac{l_1}{q},\cdots,\frac{l_s}{q}\Big) - \frac{l_1\cdots l_s}{q^s}\right| \\
&\qquad \leqslant \Big(\prod_{i=1}^{s}\frac{r_i \ln r_i n}{2\ln r_i}\Big)n^{-1}.
\end{aligned}$$

*Proof.* We may suppose that $s = 2$ as in the proof of theorem 4.4. Let

$$1 - \beta = 0.b_0'b_1'\cdots b_L'\cdots.$$

Then

$$b_v + b_v' = t - 1, \quad v = 0, 1, \cdots.$$

Hence from (4.4), we have

$$\begin{aligned}
&|N_n(\alpha,\beta) - \alpha\beta n| + |N_n(1-\alpha,\beta) - (1-\alpha)\beta n| \\
&\quad + |N_n(\alpha, 1-\beta) - a(1-\beta)n| \\
&\quad + |N_n(1-\alpha, 1-\beta) - (1-\alpha)(1-\beta)n| \\
&\leqslant ((r-1)(M+1)+2)((t-1)(L+1)+2) \\
&\quad + 2n\sum_{m=M+2}^{\infty}\frac{r-1}{r^m} + 2n\sum_{l=L+2}^{\infty}\frac{t-1}{t^l} + \frac{4}{n} \\
&\leqslant ((r-1)(M+1)+2)((t-1)(L+1)+2) + 5 \\
&\leqslant ((r-1)(M+1)+3)((t-1)(L+1)+3)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{q^2}\sum_{l=1}^{q-1}\sum_{m=1}^{q-1}\left|\frac{1}{n}N_n\Big(\frac{l}{q},\frac{m}{q}\Big)-\frac{lm}{q^2}\right|+\frac{1}{q}\sum_{l=1}^{q-1}\left|\frac{1}{n}N_n\Big(\frac{l}{q},1\Big)-\frac{l}{q}\right|\\
&\quad+\frac{1}{q}\sum_{m=1}^{q-1}\left|\frac{1}{n}N_n\Big(1,\frac{m}{q}\Big)-\frac{m}{q}\right|\\
&\leqslant\frac{1}{4n}((r-1)(M+1)+3)((t-1)(L+1)+3)\\
&\quad+\frac{1}{2n}((r-1)(M+1)+3)+\frac{1}{2n}((t-1)(L+1)+3)\\
&\leqslant\frac{1}{4n}((r-1)(M+1)+5)((t-1)(L+1)+5)\\
&<\Big(\frac{r\ \ln\ rn}{2\ \ln\ r}\Big)\Big(\frac{t\ \ln\ tn}{2\ \ln\ t}\Big)n^{-1}.
\end{aligned}$$

The theorem is proved.

**Theorem 4.5** *Suppose that* $n>\max\ (r_1,\cdots,r_{s-1})$. *Then the set*

$$\Big(\frac{k}{n},\varphi_{r_1}(k),\cdots,\varphi_{r_{s-1}}(k)\Big),\quad 1\leqslant k\leqslant n \tag{4.5}$$

*has discrepancy*

$$D(n)\leqslant\Big(\prod_{i=1}^{s-1}\frac{r_i\ \ln\ r_in}{\ln\ r_i}\Big)n^{-1}. \tag{4.6}$$

*Proof.* Let $\gamma\in G_s$. If $n\gamma_1\leqslant\max(r_1\cdots,r_{s-1})=r$ (say), then we derive from $\frac{k}{n}<\gamma_1$ that $N_n(\gamma)\leqslant r$. Hence the left hand side of (4.6) does not exceed $\frac{r}{n}$, but the right hand side of (4.6) is greater that $\frac{r}{n}$, so we have the theorem. Now, suppose that $n\gamma_1>r$. Obviously $N_n(\gamma)$ equals the number of integers $k$ satisfying

$$\varphi_{r_1}(k)<\gamma_2,\cdots,\varphi_{r_{s-1}}(k)<\gamma_s,\quad 1\leqslant k<n\gamma_1.$$

We may also suppose that $n\gamma_1$ is an integer. Otherwise we can suppose that $m<n\gamma_1<m+1$, where $m$ is an integer. Then $N_n(\gamma)$ is unchanged, if $\gamma_1$ is replaced by $\frac{m+1}{n}$. Hence by Theorem 4.3,

$$\begin{aligned}
\left|\frac{N_n(\gamma)}{n\gamma_1}-\gamma_2\cdots\gamma_s\right|&\leqslant\gamma_1^{-1}\Big(\prod_{i=1}^{s-1}\frac{r_i\ \ln\ r_i\gamma_1n}{\ln\ r_i}\Big)n^{-1}\\
&\leqslant\gamma_1^{-1}\Big(\prod_{i=1}^{s-1}\frac{r_i\ \ln\ r_in}{\ln\ r_i}\Big)n^{-1}.
\end{aligned}$$

The theorem follows.

Similar to Theorem 4.4, we have

**Theorem 4.6** *Suppose that* $n>\max(r_1^4,\cdots,r_{s-1}^4)$. *Then under the assumption of Theorem* 4.5,

$$\frac{1}{q^s}\sum_{l_1=1}^{q}\cdots\sum_{l_s=1}^{q}q^{\delta_{l_1,q}+\cdots+\delta_{l_s,q}}\left|N_n\left(\frac{l_1}{q},\cdots,\frac{l_s}{q}\right)-\frac{l_1\cdots l_s}{q^s}\right| \leqslant\left(\prod_{i=1}^{s-1}\frac{r_i\ln r_i n}{2\ln r_i}\right)n^{-1}.$$

From Theorems 4.3 and 4.5, we know that the sets (4.3) and (4.5) have discrepancies $D(n)=O\left(\frac{(\ln n)^s}{n}\right)$ and $D(n)=O\left(\frac{(\ln n)^{s-1}}{n}\right)$ respectively. Hence they are best uniformly distributed.

*Remark.* For practical use, we may take $r_i=p_i(1\leqslant i\leqslant s)$, where $p_i$ denotes the $i$-th prime number.

## 4.3 The $p$ set

As usual, we use $p$ to denote a prime number (Cf. §1.3). The sets

$$\left(\left\{\frac{k}{p}\right\},\left\{\frac{ak}{p}\right\},\cdots,\left\{\frac{a^{s-1}k}{p}\right\}\right),\quad 1\leqslant a,k\leqslant p, \tag{4.7}$$

$$\left(\left\{\frac{k}{p^2}\right\},\left\{\frac{k^2}{p^2}\right\},\cdots,\left\{\frac{k^s}{p^2}\right\}\right),\quad 1\leqslant k\leqslant p^2, \tag{4.8}$$

and

$$\left(\left\{\frac{k}{p}\right\},\left\{\frac{k^2}{p}\right\},\cdots,\left\{\frac{k^s}{p}\right\}\right),\quad 1\leqslant k\leqslant p, \tag{4.9}$$

are called the $p$ sets. The discrepancies of these sets can be evaluated with the aid of the estimates for exponential sums (Cf. Hua Loo Keng [4.5]).

**Theorem 4.7** *The set* (4.7) *has discrepancy*

$$D(p^2)<c(s)p^{-1}(\ln p)^s. \tag{4.10}$$

**Theorem 4.8** *The set* (4.8) *has discrepancy*

$$D(p^2)<c(s)p^{-1}(\ln p)^s.$$

**Theorem 4.9** *The set* (4.9) *has discrepancy*

$$D(p)<c(s)p^{-\frac{1}{2}}(\ln p)^s.$$

**Lemma 4.5** *Let $a_0, a_1, \cdots a_s$ be a set of integers such that their great common divisor is not divisible by p. Then the number of solutions of the congruence*

$$a_s x^s + \cdots + a_1 x + a_0 \equiv 0 (\text{mod } p), \ 1 \leqslant x \leqslant p$$

*is at most s* (Cf. Hua Loo Keng [2], Chap. 2).

**Lemma 4.6** *Let*

$$S = \sum_{x-1}^{p^2} e^{2\pi i f(x)/p^2},$$

*where $f(x) = a_s x^s + \cdots + a_1 x$ and at least one of the coefficients in not divisible by p. Then*

$$|S| \leqslant (s-1)p.$$

*Proof.* Since

$$f(x+py) \equiv f(x) + f'(x)py (\text{mod } p^2)$$

and

$$S = \sum_{x=1}^{p} \sum_{y=1}^{p} e^{2\pi i f(x+py)/p^2} = \sum_{x=1}^{p} e^{2\pi i f(x)/p^2} \sum_{y=1}^{p} e^{2\pi i f'(x)y/p},$$

so from Lemmas 3.6 and 4.5,

$$|S| \leqslant \sum_{x=1}^{p} \Big| \sum_{y=1}^{p} e^{2\pi i f'(x)y/p} \Big| = p \sum_{\substack{x=1 \\ f'(x) \equiv 0 (\text{mod } p)}}^{p} 1 \leqslant (s-1)p.$$

The lemma is proved.

**Lemma 4.7** *Under the assumptions of Lemma* 4.6,

$$\Big| \sum_{x=1}^{p} e^{2\pi i f(x)/p} \Big| \leqslant (s-1)\sqrt{p}$$

(Cf. A. Weil [1]).

The proof of Theorem 4.7. Let $\mathbf{a} = (1, a, \cdots, a^{s-1})$ and $\mathbf{m} = (m_1 \cdots, m_s)$. (Notice that we also use $(m_1 \cdots, m_s)$ to denote the great common divisor of $m_1, \cdots, m_s$. Please don't be confused with the notation for a vector). By Lemmas 3.6 and 4.5,

$$\sum(\mathbf{m}) = \frac{1}{p^2} \sum_{a=1}^{p} \sum_{k=1}^{p} e^{2\pi i (\mathbf{a}, \mathbf{m})k/p} = p^{-1} \sum_{\substack{a=1 \\ (\mathbf{a}, \mathbf{m}) \equiv 0 (\text{mod } p)}}^{p} 1$$

$$\leqslant \begin{cases} 1, & \text{if} \quad p | (m_1, \cdots, m_s), \\ \dfrac{s-1}{p}, & \text{if} \quad p \nmid (m_1, \cdots, m_s). \end{cases}$$

Hence by Lemma 3.2,

$$\sum_{|m_i|\leqslant p^2}' \frac{1}{||\pi \mathbf{m}||}|\sum(\mathbf{m})| \leqslant \sum_{\substack{|m_i|\leqslant p^2\\ p|(m_1,\cdots m_s)|}}' \frac{1}{||\pi \mathbf{m}||} + \frac{s-1}{p} \sum_{\substack{|m_i|\leqslant p^2\\ p\nmid(m_1,\cdots m_s)|}}' \frac{1}{||\pi \mathbf{m}||}$$

$$\leqslant \frac{s}{p}\Big(1+\frac{2}{\pi}\sum_{m=1}^{p^2}\frac{1}{m}\Big)^s < c(s)p^{-1}(\ln p)^s.$$

Take $r=1, \eta=p^{-1}$ and $h=p^2$ in Theorem 3.1 for $p>6$. Then we have the Theorem. The theorem is obvious for $p\leqslant 6$.

The proof of Theorem 4.8. Let $\mathbf{m}=(k,k^2,\cdots,k^s)$. Then from Lemma 4.6,

$$\sum_{|m_i|\leqslant p^2}' \frac{1}{||\pi \mathbf{m}||}\Big|\frac{1}{p^2}\sum_{k=1}^{p^2} e^{2\pi i(\mathbf{k},\mathbf{m})/p^2}\Big| \leqslant \sum_{\substack{|m_i|\leqslant p^2\\ p|(m_1,\cdots m_s)|}}' \frac{1}{||\pi \mathbf{m}||}$$

$$+\frac{s-1}{p}\sum_{\substack{|m_i|\leqslant p^2\\ p\nmid(m_1,\cdots m_s)|}}' \frac{1}{||\pi \mathbf{m}||} < c(s)p^{-1}(\ln p)^s.$$

Put $r=1,\eta=p^{-1}$ and $h=p^2$ in Theorem 3.1 for $p>6$. The theorem follows. For the case $p\leqslant 6$, the theorem is obvious.

The proof of Theorem 4.9. Form Lemma 4.7,

$$\sum_{|m_i|<p}' \frac{1}{||\pi \mathbf{m}||}\Big|\frac{1}{p}\sum_{k=1}^{p} e^{2\pi i(\mathbf{k},\mathbf{m})/p}\Big| \leqslant \frac{s-1}{\sqrt{p}}\sum_{|m_i|<p}' \frac{1}{||\pi \mathbf{m}||}$$

$$\leqslant \frac{s-1}{\sqrt{p}}\Big(1+\frac{2}{\pi}\sum_{m=1}^{p-1}\frac{1}{m}\Big)^s < c(s)p^{-\frac{1}{2}}(\ln p)^s.$$

Put $r=1, \eta=p^{-\frac{1}{2}}$ and $h=p-1$ in Theorem 3.1 for $p\geqslant 37$. Then we have the theorem. Form the case $p<37$, the theorem is obvious.

Now we study the lower estimate for the discrepancy of the set (4.7). Let $p\geqslant 3$. Put $x_1=1, x_2=\dfrac{2}{p}$ and $x_3=\cdots=x_s=1$. Since the congruences

$$ak\equiv 0 \text{ or } 1 (\text{mod } p), \quad 1\leqslant a,k\leqslant p$$

have $3p-2$ pairs of solutions $a,k$,

$$N_{p^2}\Big(1,\frac{2}{p},1,\cdots,1\Big)=3p-2$$

and

$$\Big|\frac{1}{p^2}N_{p^2}\Big(1,\frac{2}{p},1,\cdots,1\Big)-\frac{2}{p}\Big| = \frac{3p-2}{p^2}-\frac{2}{p}=\frac{p-2}{p^2}>\frac{1}{2p}.$$

And so the discrepancy of the set (4.7) satisfies $D(p^2) > \frac{1}{2p}$, i.e., the factor $p^{-1}$ in the right hand side of (4.10) does not admit further improvement.

*Remark.* For any given $\varepsilon > 0$, take $\eta = p^{-1}, h = [p^{1+r^{-1}}]$ and $r = r(\varepsilon)$ sufficiently large in Theorem 3.1 Then the discrepancies of the sets (4.7) and (4.8) may be given by

$$D(p^2) < \left(\left(\frac{2}{\pi}\right)^s s + \varepsilon\right)p^{-1}(\ln p)^s$$

for $p > c(s, \varepsilon)$. And if we take $\eta = p^{-\frac{1}{2}}, h = [p^{\frac{1}{2}+\frac{1}{r}}]$ and $r = r(\varepsilon)$ sufficiently large in Theorem 3.1, then the discrepancy of the set (4.9) may be given by

$$D(p) < \left(\left(\frac{1}{\pi}\right)^s (s-1) + \varepsilon\right)p^{-\frac{1}{2}}(\ln p)^s$$

for $p > c(s, \varepsilon)$.

## 4.4 The *gp* set

Let $\gamma \in G_s$. If the set of the form

$$P(k) = (\{\gamma_1 k\}, \cdots, \{\gamma_s k\}), \quad 1 \leqslant k \leqslant n \tag{4.11}$$

has discrepancy

$$D(n) \leqslant c(\gamma, \varepsilon)n^{-1+\varepsilon}, \tag{4.12}$$

then the set (4.11) is called a *gp* set and $\gamma$ a good point. Hence the sequence of sets so obtained is best uniformly distributed.

We know from Theorem 3.3 that if

$$\langle(\gamma, \mathbf{m})\rangle \geqslant c(\gamma, \varepsilon)||\mathbf{m}||^{-1-\varepsilon} \tag{4.13}$$

holds for any integral vector $\mathbf{m} \neq \mathbf{0}$, then $\gamma$ is a good point and (4.11) is a *gp* set. Hence the problem of the existence and construction of a *gp* set is reduced to the problem of the existence and construction of a vector $\gamma$ satisfying (4.13).

**Theorem 4.10** *The measure of the points $\gamma \in G_s$ such that* (4.13) *holds for any integral vector $m \neq 0$ is equal to* 1.

We derive immediately

**Theorem 4.11** *The measure of the points $\gamma \in G_s$* such that the set (4.11) *has discrepancy* (4.12) *equals* 1.

To proof Theorem 4.10, we shall need

**Lemma 4.8** *Suppose that $m$ is a non-zero integer and $c$ is a real number. Then the measure of the points satisfying*

$$\langle c+mx\rangle \leqslant \varepsilon, \quad x\in[0,1]$$

*is* $\leqslant 2\varepsilon$.

*Proof.* Clearly, we may suppose that $\varepsilon<\frac{1}{2}$ and $m>0$.

Suppose that $c=0$. Then the set of points such that

$$\langle mx\rangle \leqslant \varepsilon$$

holds, is given by the intervals

$$0\leqslant x\leqslant \frac{\varepsilon}{m},\quad \frac{1-\varepsilon}{m}\leqslant x\leqslant \frac{1+\varepsilon}{m},\quad \frac{2-\varepsilon}{m}\leqslant x\leqslant \frac{2+\varepsilon}{m},\cdots,$$
$$\frac{m-1-\varepsilon}{m}\leqslant x\leqslant \frac{m-1+\varepsilon}{m},\quad \frac{m-\varepsilon}{m}\leqslant x\leqslant 1.$$

So its measure is$\leqslant 2\varepsilon$.

Suppose that $c\neq 0$. Then

$$c+mx=m\Big(x+\frac{c}{m}\Big)=my,$$

so it is reduced to the case $c=0$. The lemma follows.

Theorem 4.10 is a consequence of the following

**Lemma 4.9** *If $\psi(\bar{n})>0$ and*

$$\sum_{n=-\infty}\frac{1}{\bar{n}\psi(\bar{n})}<\infty,$$

*then the inequality*

$$\langle(\boldsymbol{\gamma},\mathbf{m})\rangle > \frac{c(\boldsymbol{\gamma},\psi)}{\prod\limits_{i=1}^{s}\bar{m}_i\psi(\bar{m}_i)},\quad \mathbf{m}\neq\mathbf{0}$$

*is satisfied for almost all $\boldsymbol{\gamma}\in G_s$.*

*Proof.* We may easily prove by Lemma 4.8 and mathematical induction that the measure of points $\boldsymbol{\gamma}$ satisfying

$$\langle(\boldsymbol{\gamma},\mathbf{m})\rangle \leqslant \varepsilon$$

is $\leqslant 2\varepsilon$, if $\mathbf{m} \neq \mathbf{0}$. Hence the measure of the set $\sigma_\eta$ of point $\gamma$ such that the inequality

$$\langle(\gamma, \mathbf{m})\rangle \leqslant \frac{\eta}{\prod_{i=1}^{s} \bar{m}_i \psi(\bar{m}_i)}, \quad \eta > 0$$

holds for any $\mathbf{m} \neq \mathbf{0}$ does not exceed

$$2\eta \sum{}' \frac{1}{\prod_{i=1}^{s} \bar{m}_i \psi(\bar{m}_i)} = 2\eta\left(\Big(\sum_{m=-\infty}^{\infty} \frac{1}{\bar{m}\psi(\bar{m})}\Big)^s - \frac{1}{\psi(1)^s}\right)$$
$$= c(\psi)\eta(\text{say}).$$

Now we shall prove that for any $\tau > 0$, the measure of the set $\sigma$ of $\gamma$ such that the inequality

$$\langle(\gamma, \mathbf{m})\rangle > \frac{\tau}{\prod_{i=1}^{s} \bar{m}_i \psi(\bar{m}_i)}$$

can not be satisfied by all $\mathbf{m} \neq \mathbf{0}$ is equal to zero. Otherwise, suppose that $\sigma$ has the measure $\delta > 0$. Then for any $\eta > 0, \sigma$ is the subset of $\sigma_\eta$. Take $\eta = \dfrac{\delta}{2c(\psi)}$. Then the measure of $\sigma$ does not exceed the measure of $\sigma_\eta$i.e.,$\dfrac{\delta}{2}$. This gives a contradiction. The lemma is proved.

## 4.5 The construction of good points

Theorem 4.11 means that the measure of the set of good points in $G_s$ equals 1, but Theorem 4.11 is not constructive. Two constructive results were obtained by W. M. Schmidt and A. Baker respectively.

**Theorem 4.12** *Let $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_s)$, where $\alpha_i (1 \leqslant i \leqslant s)$ is a set of real algebraic numbers such that* 1, $\alpha_1 \cdots, \alpha_s$ *are linearly independent over Q. Then*

$$\langle(\boldsymbol{\alpha}, \mathbf{m})\rangle > c(\boldsymbol{\alpha},\varepsilon)||\mathbf{m}||^{-1-\varepsilon}$$

*holds for any integral vector* $\mathbf{m} \neq \mathbf{0}$ (Cf. W. M. Schmidt [2.4]).

**Theorem 4.13** *Let $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_s)$, where $\beta_i = e^{r_i} (1 \leqslant i \leqslant s)$ with $r_i (1 \leqslant i \leqslant s)$ denoting a set of different non-zero rational numbers. Then*

$$\langle(\boldsymbol{\beta}, \mathbf{m})\rangle > c(\boldsymbol{\beta},\varepsilon)||\mathbf{m}||^{-1-\varepsilon}$$

*holds for any integral vector* $\mathbf{m} \neq \mathbf{0}$ (Cf. A. Barker[1]).

From Theorem 4.12 and 4.13, we derive immediately

**Theorem 4.14** *The set*

$$P(k) = (\{a_1 k\}, \cdots, \{a_s k\}), \quad 1 \leqslant k \leqslant n$$

*has discrepancy*

$$D(n) \leqslant c(\boldsymbol{\alpha}, \varepsilon) n^{-1+\varepsilon}.$$

**Theorem 4.15** *The set*

$$P(k) = (\{\beta_1 k\}, \cdots, \{\beta_s k\}), \quad 1 \leqslant k \leqslant n$$

*has discrepancy*

$$D(n) \leqslant c(\boldsymbol{\beta}, \varepsilon) n^{-1+\varepsilon}.$$

*Remark.* The constant in Theorem 4.13 is effective (Cf. K. Mahler [1]).

## 4.6 The $\mathscr{R}_s$ set

Let $h_1, \cdots, h_s, n(>0)$ be a set of integers. If the set of rational points

$$\left(\left\{\frac{h_1 k}{n}\right\}, \cdots \left\{\frac{h_s k}{n}\right\}\right), \quad 1 \leqslant k \leqslant n \tag{4.14}$$

has discrepancy

$$D(n) \leqslant c(s, \varepsilon) n^{-1+\varepsilon},$$

the set (4.14) is called a *glp* set and $\mathbf{h}$ an optimal coefficient mod $n$ or a good lattice point mod $n$.

Notice that the definition of optimal coefficient or good lattice point given here is different from the original definition due to Korobov and Hlawka (Cf. N. M. Kopobov [2] and E. Hlawka [3]).

Let $\mathscr{R}_s$ be a real algebraic number field of degree $s$ and $\omega_1, \cdots, \omega_s$ be an integral basis of $\mathscr{R}_s$, where $\omega_2, \cdots, \omega_s$ are irrational numbers . Let $(n_l, h_{l1}, \cdots, h_{ls})(l = 1, 2, \cdots)$ be a sequence of sets of integers satisfying

$$\left|\frac{h_{lj}}{n_l} - \omega_j\right| \leqslant c(\mathscr{R}_s) n_l^{-1-\frac{1}{s-1}}, \quad 2 \leqslant j \leqslant s. \tag{4.15}$$

For simplicity, we omit the index $l$.

**Theorem 4.16** *The set*

$$\left(\left\{\frac{k}{n}\right\},\left\{\frac{h_2k}{n}\right\},\cdots,\left\{\frac{h_sk}{n}\right\}\right),\quad 1\leqslant k\leqslant n \tag{4.16}$$

*has discrepancy*

$$D(n)\leqslant c(\mathscr{R}_s,\varepsilon)n^{-\frac{1}{2}-\frac{1}{2(s-1)+\varepsilon}}.$$

**Theorem 4.17** *Suppose that* $1\leqslant q\leqslant n^{\frac{1}{2}+\frac{1}{2(s-1)}}$. *Then the set*

$$\left(\left\{\frac{h_2k}{n}\right\},\cdots,\left\{\frac{h_sk}{n}\right\}\right),\quad 1\leqslant k\leqslant q \tag{4.17}$$

*has discrepancy*

$$D(q)\leqslant c(\mathscr{R}_s,\varepsilon)q^{-1+\varepsilon}.$$

Since $1,\omega_2,\cdots,\omega_s$ is a basis of $\mathscr{R}_s$, they are linearly independent over the rational field Q. Hence Theorems 4.16 and 4.17 follow by (4.15) and Theorems 3.4, 3.5 and 4.12.

The sets (4.16) and (4.17) are called the $\mathscr{R}_s$ sets, Especially, for the real cyclotomic field $\mathscr{R}_s=Q\left(\cos\frac{2\pi}{m}\right)$, where $s=\varphi(m)/2$, we have

$$\left|\frac{c_j}{n}-2\cos\frac{2\pi(j-1)}{m}\right|<c(\mathscr{R}_s,\varepsilon)n^{-1-\frac{1}{s-1}},\quad 2\leqslant j\leqslant s$$

(Cf.§1.3) and so the following

**Theorem 4.18** *The set*

$$\left(\left\{\frac{c_1k}{n}\right\},\cdots,\left\{\frac{c_sk}{n}\right\},\quad 1\leqslant k\leqslant n\right.$$

*has discrepancy*

$$D(n)\leqslant c(\mathscr{R}_s,\varepsilon)n^{-\frac{1}{2}-\frac{1}{2(s-1)}+\varepsilon}.$$

**Theorem 4.19** *Suppose that* $1\leqslant q\leqslant n^{\frac{1}{2}+\frac{1}{2(s-1)}}$. *Then the set*

$$\left(\left\{\frac{c_2k}{n}\right\},\cdots,\left\{\frac{c_sk}{n}\right\}\right),\quad 1\leqslant k\leqslant q$$

*has discrepancy*

$$D(q)\leqslant c(\mathscr{R}_s,\varepsilon)q^{-1+\varepsilon}.$$

## 4.7 The $\eta$ set

Let $\alpha$ be a PV number of degree $s$, i.e., $\alpha > 1$ and its conjugates satisfy

$$|\alpha^{(2)}| \leqslant \cdots \leqslant |\alpha^{(s)}| < 1.$$

Let

$$\rho = -\frac{\ln|\alpha^{(s)}|}{\ln \alpha}.$$

Let $\alpha$ satisfy the irreducible polynomial

$$x^s - a_{s-1}x^{s-1} - \cdots - a_1 x - a_0 = 0$$

and $Q_n (n = 0, 1, \cdots)$ be a set of integers defined by the recurrence relation

$$\begin{aligned} &Q_0 = Q_1 = \cdots = Q_{s-2} = 0, \quad Q_{s-1} = 1, \\ &Q_n = a_{s-1}Q_{n-1} + \cdots + a_1 Q_{n-s+1} + a_0 Q_{n-s}, \quad n \geqslant s. \end{aligned}$$

Further let

$$Q_n(j) = Q_{n+j-1} - a_{s-1}Q_{n+j-2} - \cdots - a_{s-j+2}Q_{n+1} - a_{s-j+1}Q_n, \quad 2 \leqslant j \leqslant s.$$

Then

$$\left|\frac{Q_n(j)}{Q_n} - \omega_j\right| < c(\alpha)|Q_n|^{-1-\rho}, \quad 2 \leqslant j \leqslant s, n > c_1(\alpha) \tag{4.18}$$

where

$$\omega_j = \alpha^{j-1} - a_{s-1}\alpha^{j-2} - \cdots - a_{s-j+2}\alpha - a_{s-j+1} \quad (2 \leqslant j \leqslant s)(\text{Cf. §2.2}).$$

**Theorem 4.20** *The set*

$$\left(\left\{\frac{k}{Q_n}\right\}, \left\{\frac{Q_n(2)}{Q_n}k\right\}, \cdots, \left\{\frac{Q_n(s)}{Q_n}k\right\}\right) \quad 1 \leqslant k \leqslant |Q_n| \tag{4.19}$$

*has discrepancy*

$$D(Q_n) \leqslant c(\alpha, \varepsilon)|Q_n|^{-\frac{1}{2}-\frac{\rho}{2}+\varepsilon}.$$

**Theorem 4.21** *Suppose that* $1 \leqslant q \leqslant |Q_n|^{\frac{1}{2}+\frac{\rho}{2}}$. *Then the set*

$$\left(\left\{\frac{Q_n(2)}{Q_n}k\right\}, \cdots, \left\{\frac{Q_n(s)}{Q_n}k\right\}\right), \quad 1 \leqslant k \leqslant q \tag{4.20}$$

*has discrepancy*

$$D(q) \leqslant c(\alpha, \varepsilon)q^{-1+\varepsilon}.$$

Since $1, \omega_2, \cdots, \omega_s$ form a basis of $Q(\alpha)$, hence Theorems 4.20 and 4.21 follow by (4.18), and Theorems 3.4, 3.5 and 4.12.

The sets (4.19) and (4.20) are called the $\alpha$ set. Especially, we take the $\eta$ set, where $\eta$ is the greatest real root of the equation

$$x^s - x^{s-1} - \cdots - x - 1 = 0.$$

Let $F_n (= F_{s,n})$ be a set of integers defined by the recurrence relation

$$\begin{aligned} &F_0 = F_1 = \cdots = F_{s-2} = 0, \quad F_{s-1} = 1, \\ &F_n = F_{n-1} + \cdots + F_{n-s+1} + F_{n-s}, \quad n \geqslant s, \end{aligned}$$

Further let

$$F_n(j) = F_{n+j-1} - F_{n+j-2} - \cdots - F_n, \quad 2 \leqslant j \leqslant s.$$

Then

$$\left| \frac{F_n(j)}{F_n} - \omega_j \right| \leqslant c(\eta) F_n^{-1-\frac{1}{2^s \ln 2}-\frac{1}{2^{2s+1}}}, \quad 2 \leqslant j \leqslant s,$$

where $\omega_j = \eta^{j-1} - \eta^{j-2} - \cdots - \eta - 1$(Cf. §2.8). Hence we have

**Theorem 4.22** *The set*

$$\left( \left\{ \frac{k}{F_n} \right\}, \left\{ \frac{F_n(2)}{F_n} k \right\}, \cdots, \left\{ \frac{F_n(s)}{F_n} k \right\} \right) \quad 1 \leqslant k \leqslant F_n \tag{4.21}$$

*has discrepancy*

$$D(F_n) \leqslant c(\eta) F_n^{-\frac{1}{2}-\frac{1}{2^{s+1}\ln 2}-\frac{1}{2^{2s+3}}}. \tag{4.22}$$

**Theorem 4.23** *Suppose that* $1 \leqslant q \leqslant F_n^{\frac{1}{2}+\frac{1}{2^{s+1}\ln 2}+\frac{1}{2^{2s+2}}}$. *Then the set*

$$\left( \left\{ \frac{F_n(2)}{F_n} k \right\} \cdots, \left\{ \frac{F_n(s)}{F_n} k \right\} \right), \quad 1 \leqslant k \leqslant q$$

*has discrepancy*

$$D(q) \leqslant c(\eta, \varepsilon) q^{-1+\varepsilon}.$$

For real quadratic irrational $\alpha$, we have

$$\langle \alpha m \rangle > \frac{c(\alpha)}{|m|} \tag{4.23}$$

(Cf. Hua Loo Keng [2], Chap, 10). If we use (4.23) and Theorem 2.9 to instead of Theorem 4.12 and Theorem 2.8 respectively, then we have

**Theorem 4.24** *For* $s = 2$, *the right hand side of* (4.22) *may be replaced by* $c(\eta) F_n^{-1} (\ln F_n)^2$.

Hence the set (4.21) is a *glp* set for $s = 2$. For $s = 3$, we may use Theorem 2.9 instead of theorem 2.8. Then we have

**Theorem 4.25** *For $s = 3$, the right hand side of (4.22) and the range of $q$ in Theorem 4.23 may be replaced by $c(\eta, \varepsilon)F_n^{-3/4+\varepsilon}$ and $q \leqslant F_n^{3/4}$ respectively.*

## 4.8 The case $s$=2

Let

$$a_3, a_4, \cdots$$

be a set of positive integers such that $a_n \leqslant M(n = 3, 4, \cdots)$. Let $q_1, q_2$ be two positive integers such that $q_1 \leqslant q_2$ and $(q_1, q_2) = 1$. Further let

$$q_n = a_n q_{n-1} + q_{n-2}, \quad n \geqslant 3. \tag{4.24}$$

Then $q_n (n = 1, 2, \cdots)$ form an increasing sequence of integers and

$$n - 1 \leqslant q_n \leqslant (M + 1)q_{n-1}, \quad n \geqslant 3. \tag{4.25}$$

**Theorem 4.26** *There exists a constant $c(q_1, q_2, M)$ such that the solution of the equation*

$$\begin{aligned} &x_1 + q_{n-1}x_2 = q_n y, \quad 0 < |x_1| \leqslant q_n/2, \\ &0 < |x_2| \leqslant q_n/2, \quad y \neq 0 \end{aligned} \tag{4.26}$$

*satisfies*

$$|x_1 x_2| \geqslant c q_n, \quad |x_1 y| \geqslant c q_{n-1}.$$

*Proof.* The theorem holds for $n = 2, 3$ obviously. Suppose that $n > 3$ and that the theorem holds for any integer$< n$. Now we proceed to prove that the theorem holds for $n$ also.

Clearly, $x_2$ and $y$ have the same sign. Otherwise

$$\frac{q_n}{2} \geqslant |x_1| = |q_n y - q_{n-1}x_2| \geqslant q_n + q_{n-1}.$$

This leads to a contradiction.

If $|x_1| \leqslant \frac{1}{2}q_{n-1}, |y| \leqslant \frac{1}{2}q_{n-1}$, then by (4.24) and (4.26),

$$\begin{aligned} &x_1 + q_{n-1}x_2 = y(a_n q_{n-1} + q_{n-2}), \\ &x_1 - q_{n-2}y = q_{n-1}(a_n y - x_2). \end{aligned} \tag{4.27}$$

If $a_n y = x_2$, then by (4.27), $x_1 = q_{n-2}y$, hence from (4.25),

$$x_1x_2 = a_nq_{n-2}y^2 \geqslant q_{n-2} \geqslant \frac{1}{M+1}q_{n-1} \geqslant \frac{1}{(M+1)^2}q_n,$$
$$x_1y = q_{n-2}y^2 \geqslant q_{n-2} \geqslant \frac{1}{M+1}q_{n-1}.$$

Take $c \leqslant \frac{1}{(M+1)^2}$. The theorem follows.

Now, suppose that $a_n y \neq x_2$. Then it follows from the induction hypothesis that there exists a constat $c = c(q_1, q_2, M)$ such that the solution of the equation (4.26) satisfies

$$|x_1y| \geqslant cq_{n-1}, \quad |x_1(a_ny - x_2)| \geqslant cq_{n-2}. \tag{4.28}$$

Since $y$ and $x_2 - a_ny$ have the same sign, we have

$$|x_1x_2| = |x_1(x_2 - a_ny) + x_1a_ny| \geqslant cq_{n-2} + ca_nq_{n-1} = cq_n$$

by (4.28).

If $|x_1| \geqslant \frac{1}{2}q_{n-1}$, then by (4.25),

$$|x_1x_2| \geqslant \frac{1}{2}q_{n-1} \geqslant \frac{1}{2(M+1)}q_n,$$
$$|x_1y| \geqslant \frac{1}{2}q_{n-1}.$$

If $|y| \geqslant \frac{1}{2}q_{n-1}$, then

$$|x_1y| \geqslant \frac{1}{2}q_{n-1},$$
$$\frac{1}{2}q_{n-1}q_n \leqslant |y|q_n \leqslant |x_1| + q_{n-1}|x_2| \leqslant \frac{1}{2}q_n + q_{n-1}|x_2|,$$
$$|x_2| \geqslant \frac{1}{2}q_n\left(1 - \frac{1}{q_{n-1}}\right) \geqslant \frac{1}{4}q_n.$$

So

$$|x_1x_2| \geqslant \frac{1}{4}q_n.$$

Hence the theorem holds for $n$ too. The theorem follows by induction.

**Theorem 4.27** *There exists a constant $c = c(q_1, q_2, M)$ such that the congruence*

$$x_1 + q_{n-1}x_2 \equiv 0 (\text{mod } q_n) \tag{4.29}$$

*has no solution in the domain*

$$\bar{x}_1\bar{x}_2 < cq_n, \quad (x_1, x_2) \neq (0,0).$$

*Proof.* Since $(q_1, q_2) = 1$, we have $(q_{n-1}, q_n) = 1 (n = 3, 4, \cdots)$ by (4.24). If $(x_1, x_2) \neq (0,0)$ is a solution of (4.29), then $q_n | x_2$ if $x_1 = 0$ and $q_n | x_1$ if $x_2 = 0$. Hence $\bar{x}_1\bar{x}_2 \geqslant q_n$. If $x_1 \neq 0$ and $x_2 \neq 0$, then it follows by Theorem 4.26 that the solution of (4.29) satisfies $\bar{x}_1\bar{x}_2 \geqslant cq_n$. The theorem is proved.

From Theorem 3.2, we derive

**Theorem 4.28** . *The set*

$$\left(\left\{\frac{k}{q_n}\right\}, \left\{\frac{q_{n-1}k}{q_n}\right\}\right), \quad 1 \leqslant k \leqslant q_n$$

*has discrepancy*

$$D(q_n) \leqslant c(q_1, q_2, M) q_n^{-1} (\ln 3q_n)^2.$$

Take $q_1 = q_2 = 1$ and $a_n = 1 (n = 3, 4, \cdots)$. Then the sequence $q_n (n = 1, 2, \cdots)$ is the usual Fibonacci sequence $F (= F_{2,n}) (n = 1, 2, \cdots)$ of dimension 2. Hence Theorem 4.24 may be derived by Theorem 4.28 also.

## 4.9 The *glp* set

**Theorem 4.29** *There exists an integral vector* $\mathbf{a} (= \mathbf{a}(p))$ *such that the set*

$$\left(\left\{\frac{a_1 k}{p}\right\}, \cdots, \left\{\frac{a_s k}{p}\right\}\right), \quad 1 \leqslant k \leqslant p \tag{4.30}$$

*has discrepancy*

$$D(p) < c(s) p^{-1} (\ln p)^s.$$

**Theorem 4.30** *Suppose that* $n$ *is an integer satisfying* $1 \leqslant n \leqslant p$. *Then there exists an integral vector* $\mathbf{a} (= \mathbf{a}(p))$ *such that the set*

$$\left(\left\{\frac{a_1 k}{p}\right\}, \cdots, \left\{\frac{a_s k}{p}\right\}\right), \quad 0 \leqslant k < n \tag{4.31}$$

*has discrepancy*

$$D(n) < c(s) n^{-1} (\ln p)^{s+1}.$$

To prove these theorems, we shall need

**Lemma 4.10** *Let* $\mathbf{a}$ *be an integral vector and* $q$ *an integer* $> 1$. *If* $(a_i, q) = 1 (1 \leqslant i \leqslant s)$, *then for any positive integer* $r$,

$$\sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0 (\bmod\ q)\\ |m_i|\leqslant qr}}{}' \frac{1}{\|\pi\mathbf{m}\|} - \sum_{\substack{(\mathbf{a},\mathbf{m}^{(0)})\equiv 0(\bmod\ q)\\ -q/2<m_i^{(0)}\leqslant q/2}}{}' \frac{1}{\|\pi\mathbf{m}^{(0)}\|} < \frac{s2^s(\ln 20qr)^s}{\pi^s q}.$$

*Proof.* If $\mathbf{m}^{(0)}$ is a solution of

$$(\mathbf{a}, \mathbf{m}^{(0)}) = a_1 m_1^{(0)} + \cdots + a_s m_s^{(0)} \equiv 0 (\bmod\ q)$$
$$-q/2 < m_i^{(0)} \leqslant q/2, \quad 1 \leqslant i \leqslant s.$$

Then

$$m_1 = m_1^{(0)} + ql_1, \cdots, m_s = m_s^{(0)} + ql_s \tag{4.32}$$

is also a solution of the congruence

$$(\mathbf{a}, \mathbf{m}) \equiv 0 (\bmod\ q). \tag{4.33}$$

On the other hand, any solution of (4.33) may be represented by (4.32). If and $s-1$ integers of $m_1^{(0)}, \cdots, m_s^{(0)}$ are given, then the remaining one can be determined uniquely, since $(a_i, q) = 1 (1 \leqslant i \leqslant s)$. By Lemma 3.2, we have

$$\sum_{-q/2<m\leqslant q/2} \sum_{l=-r}^{r} \frac{1}{\overline{\pi(m+lq)}} \leqslant 1 + \frac{2}{\pi} \sum_{m=1}^{[q(r+\frac{1}{2})]} \frac{1}{m} < \frac{2}{\pi} \ln 20qr$$

and

$$\sum_{l=-r}^{r}{}' \frac{1}{\overline{\pi(m+lq)}} \leqslant \frac{2}{\pi} \sum_{l=1}^{r} \frac{1}{q\left(l-\frac{1}{2}\right)} < \frac{2}{\pi q} \ln 20r$$

for $-q/2 < m \leqslant q/2$. Hence

$$\left| \sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0 (\bmod\ q)\\ |m_i|\leqslant qr}}{}' \frac{1}{\|\pi\mathbf{m}\|} - \sum_{\substack{(\mathbf{a},\mathbf{m}^{(0)})\equiv 0(\bmod\ q)\\ -q/2<m_i^{(0)}\leqslant q/2}}{}' \frac{1}{\|\pi\mathbf{m}^{(0)}\|} \right|$$
$$\leqslant \sum_{v=1}^{s} \sum_{\substack{(\mathbf{a},\mathbf{m}^{(0)})\equiv 0(\bmod\ q)\\ -q/2<m_v^{(0)}\leqslant q/2}} \Big( \sum_{l_v=-r}^{r}{}' \frac{1}{\overline{\pi(m_v^{(0)}+l_v q)}} \Big) \prod_{\substack{\mu=1\\ \mu\neq v}}^{s} \Big( \sum_{l_\mu=-r}^{r} \frac{1}{\overline{\pi(m_\mu^{(0)}+l_\mu q)}} \Big)$$
$$< s\left(\frac{2}{\pi}\right)^s q^{-1} (\ln 20qr)^s.$$

The lemma is proved.

**Lemma 4.11** *There exists an integral vector* $\mathbf{a}(=\mathbf{a}(p))$ *such that*

$$\sideset{}{'}\sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0 (\text{mod } p)\\ |m_i|<p/2}} \frac{1}{||\pi\mathbf{m}||} < c(s)p^{-1}(\ln p)^s.$$

*Proof.* Let $\mathbf{a}=(1,a,\cdots,a^{s-1})$, where $a$ is an integer. Let

$$\Lambda(a)=\sideset{}{'}\sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0 (\text{mod } p)\\ |m_i|<p/2}} \frac{1}{||\pi\mathbf{m}||}.$$

Then by lemma 4.5,

$$\begin{aligned}\min_{1\leqslant\alpha\leqslant p}\Lambda(a) &\leqslant \frac{1}{p}\sum_{\alpha=1}^{p}\Lambda(a)=p^{-1}\sideset{}{'}\sum_{|m_i|<p/2}\frac{1}{||\pi\mathbf{m}||}\sum_{\substack{1\leqslant\alpha\leqslant p\\(\mathbf{a},\mathbf{m})\equiv 0(\text{mod } p)}}1\\ &\leqslant (s-1)p^{-1}\Big(\sum_{|m|<p/2}\frac{1}{\pi\bar{m}}\Big)^s < c(s)p^{-1}(\ln p)^s.\end{aligned}$$

Hence there exists an integer $a$ such that

$$\Lambda(a) < c(s)p^{-1}(\ln p)^s.$$

The lemma is proved.

The proof of Theorem 4.29. Take $\mathbf{a}$ satisfying Lemma 4.11. Then by Lemmas 4.10 and 4.11,

$$\begin{aligned}\sideset{}{'}\sum_{|m_i|\leqslant p^2}\frac{1}{||\pi\mathbf{m}||}\Big|\frac{1}{p}\sum_{k=1}^{p}e^{2\pi i(\mathbf{a},\mathbf{m})k/p}\Big| &= \sideset{}{'}\sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0(\text{mod } p)\\ |m_i|\leqslant p^2}}\frac{1}{||\pi\mathbf{m}||}\\ &< \sideset{}{'}\sum_{\substack{(\mathbf{a},\mathbf{m}^{(0)})\equiv 0(\text{mod } p)\\ |m_i^{(0)}|<p/2}}\frac{1}{||\pi\mathbf{m}^{(0)}||}+\frac{s2^{2s}(\ln 8p)^s}{\pi^s p}\\ &< c(s)p^{-1}(\ln p)^s.\end{aligned}$$

Take $r=1,\eta=p^{-1}$ and $h=p^2$ in Theorem 3.1 for $p>6$. Then we have the result. The theorem is obvious for $p\leqslant 6$.

The proof of Theorem 4.30. Let $\mathbf{a}=(1,a,\cdots,a^s)$ be a vector such that Theorem 4.29 holds. Then for $\gamma\in G_s$, the number $N_n(\gamma)$ of integers $k$ satisfying

$$\left\{\frac{ak}{p}\right\}<\gamma_1,\cdots,\left\{\frac{a^s k}{p}\right\}<\gamma_s,\quad 0\leqslant k<n$$

is equal to the number of integers $k$ satisfying

$$\left\{\frac{k}{p}\right\} < \frac{n}{p}, \left\{\frac{ak}{p}\right\} < \gamma_1, \cdots, \left\{\frac{a^s k}{p}\right\} < \gamma_s, \quad 1 \leqslant k \leqslant p.$$

Hence from Theorem 4.29,

$$\left|p^{-1}N_n(\gamma) - \frac{n}{p}|\gamma|\right| < c(s)p^{-1}(\ln p)^{s+1}$$

the theorem follows.

The vector $\mathbf{a}$ in Theorem 4.29 is given by the integer $a$ such that $\Lambda(a)$ is minimal. Now we shall give an explicit expression for $\Lambda(a)$.

**Lemma 4.12** *Let $q$ be a positive integer and $x$ satisfy $0 < x < 1$. Then*

$$1 - \frac{2}{\pi}\ln(2\sin \pi x) = \sum_{|m|<q} \frac{e^{2\pi imx}}{\pi \bar{m}} + \frac{\psi}{\pi q\langle x\rangle}.$$

*Hereafter $\psi$ denotes a number satisfying $|\psi| \leqslant 1$.*

*Proof.* For $0 \leqslant r < 1$.

$$\sum_{m=1}^{\infty} \frac{r^m e^{2\pi imx}}{m} = -\ln(1 - re^{2\pi ix}).$$

Since the series $\sum_{m=1}^{\infty} \frac{e^{2\pi imx}}{m}$ is convergent, so

$$\sum_{m=1}^{\infty} \frac{e^{2\pi imx}}{m} = -\ln(1 - e^{2\pi ix})$$

and

$$\begin{aligned}\sum_{m=-\infty}^{\infty} \frac{e^{2\pi imx}}{\pi \bar{m}} &= 1 + \pi^{-1}\sum_{m=1}^{\infty} \frac{e^{2\pi imx}}{m} + \pi^{-1}\sum_{m=1}^{\infty} \frac{e^{-2\pi imx}}{m} \\ &= 1 - \pi^{-1}(\ln(1 - e^{2\pi ix}) + \ln(1 - e^{-2\pi ix})) \\ &= 1 - \pi^{-1}\ln(2 - 2\cos 2\pi x) = 1 - \frac{2}{\pi}\ln(2\sin \pi x).\end{aligned}$$

Hence

$$1 - \frac{2}{\pi}\ln(2\sin \pi\, x) = \sum_{|m|<q} \frac{e^{2\pi imx}}{\pi \bar{m}} + R,$$

where

$$R = \sum_{m=q}^{\infty} \frac{e^{2\pi imx}}{\pi m} + \sum_{m=q}^{\infty} \frac{e^{-2\pi imx}}{\pi m}.$$

From the identity

$$\frac{e^{2\pi imx}}{m} = \frac{1}{e^{2\pi ix}-1}\Big(\frac{e^{2\pi i(m+1)x}}{m+1} - \frac{e^{2\pi imx}}{m} + \frac{e^{2\pi i(m+1)x}}{m(m+1)}\Big),$$

we have

$$\begin{aligned}\Big|\sum_{m=q}^{\infty}\frac{e^{2\pi imx}}{\pi m}\Big| &= \frac{1}{\pi|e^{2\pi ix}-1|}\Big|-\frac{e^{2\pi iqx}}{q} + \sum_{m=q}^{\infty}\frac{e^{2\pi i(m+1)x}}{m(m+1)}\Big| \\ &\leqslant \frac{1}{2\pi\sin\pi x}\Big(\frac{1}{q} + \sum_{m=q}^{\infty}\frac{1}{m(m+1)}\Big) = \frac{1}{q\pi\sin\pi x} \\ &\leqslant \frac{1}{2\pi q\langle x\rangle}.\end{aligned}$$

Hence

$$R = -\frac{\psi}{\pi q\langle x\rangle}.$$

The lemma is proved.

Since

$$\begin{aligned}\Big|1-\frac{2}{\pi}\ln\Big(2\sin\pi\Big\{\frac{a^{v-1}k}{p}\Big\}\Big)\Big| &\leqslant 1+\frac{2}{\pi}\Big|\ln\Big(2\sin\frac{\pi}{p}\Big)\Big| \\ &\leqslant 1+\frac{2}{\pi}\ln p < \frac{2}{\pi}\ln 8p\end{aligned}$$

for $1 \leqslant a, k \leqslant p$, then by Lemma 4.12,

$$\begin{aligned}\prod_{v=1}^{s}\Big(1-\frac{2}{\pi}\ln\Big(2\sin\pi\Big\{\frac{a^{v-1}k}{p}\Big\}\Big)\Big) &= \sum_{|m_i|<p/2}\frac{e^{2\pi i(\mathbf{a},\mathbf{m})k/p}}{||\pi\mathbf{m}||} \\ &+\psi\Big(\frac{2}{\pi}\Big)^{s-1}(\ln 8p)^{s-1}\frac{2}{\pi(p+1)}\sum_{v=1}^{s}\frac{1}{\langle\frac{a^{v-1}k}{p}\rangle}.\end{aligned}$$

Since

$$\sum_{k=1}^{p-1}\frac{1}{\langle\frac{a^{v-1}k}{p}\rangle} = 2p\sum_{k=1}^{\frac{p-1}{2}}\frac{1}{k} < 2p\ln 8p$$

then, by Lemma 3.2,

$$\begin{aligned}\Lambda(a) &= p^{-1}\sum_{k=1}^{p}\sum_{|m_i|<p/2}\frac{e^{2\pi i(\mathbf{a},\mathbf{m})k/p}}{||\pi\mathbf{m}||}-1\\
&= p^{-1}\sum_{k=1}^{p-1}\sum_{|m_i|<p/2}\frac{e^{2\pi i(\mathbf{a},\mathbf{m})k/p}}{||\pi\mathbf{m}||}-1+p^{-1}\sum_{|m_i|<p/2}\frac{1}{||\pi\mathbf{m}||}\\
&= p^{-1}\sum_{k=1}^{p-1}\prod_{v=1}^{s}\Big(1-\frac{2}{\pi}\ln\Big(2\sin\pi\Big\{\frac{a^{v-1}k}{p}\Big\}\Big)\Big)\\
&\quad+\frac{\psi 2^{s+1}}{\pi^s}p^{-1}(\ln 8p)^s-1+\psi\frac{2^s}{\pi^s}p^{-1}(\ln 8p)^s\\
&= p^{-1}\sum_{k=1}^{p-1}\prod_{v=1}^{s}\Big(1-\frac{2}{\pi}\ln\Big(2\sin\pi\Big\{\frac{a^{v-1}k}{p}\Big\}\Big)\Big)\\
&\quad-1+3\psi\Big(\frac{2}{\pi}\Big)^s p^{-1}(\ln 8p)^s.\end{aligned}$$

For $a = 1, 2, \cdots, p-1$, we may find the integer $a$ such that $\Lambda(a)$ assumes a minimum or satisfies

$$\sum_{k=1}^{p-1}\prod_{v=1}^{s}\Big(1-\frac{2}{\pi}\ln\Big(2\sin\pi\Big\{\frac{a^{v-1}k}{p}\Big\}\Big)\Big)\leqslant p+O((\ln p)^s).$$

Hence we have the vector $\mathbf{a} = (1, a, \cdots, a^{s-1})$ satisfying Theorem 4.29.

The set (4.30) is a *glp* set which has not only a precise discrepancy but is also very convenient for use. However, it has the disadvantage that it requires heavy computations to obtain the integer $a$. Roughly speaking, it requires $O(p^2)$ elementary operations (Cf. N. M. Korobov [5.7]).

*Remark.* For $\varepsilon > 0$, take $\mathbf{r} = r(\varepsilon)$ sufficiently large, $\eta = p^{-1}$ and $h = [p^{1+r^{-1}}]$. Then the discrepancies of the sets (4.30) and (4.31) may be replaced by

$$D(p) < \Big(\frac{(2s-1)2^s}{\pi^s}+\varepsilon\Big)p^{-1}(\ln p)^s$$

and

$$D(n) < \Big(\frac{(2s+1)2^{s+1}}{\pi^{s+1}}+\varepsilon\Big)n^{-1}(\ln p)^{s+1}$$

respectively for $p > c(s, \varepsilon)$.

**Notes**

J. G. Van der Corput [1] first gave the best uniformly distributed set $\Big(\frac{k}{n}, \varphi_2(k)\Big)$ $(1 \leqslant k \leqslant n)$.

J. M. Hammersley [1] and J. H. Halton [1] suggested the generalizations $\left(\frac{k}{n}, \varphi_{p_1}(k), \cdots, \varphi_{p_{s-1}}(k)\right)(1 \leqslant k \leqslant n)$ and $(\varphi_{p_1}(k), \cdots, \varphi_{p_s}(k))(1 \leqslant k \leqslant n)$ respectively. Theorem 4.3 was proved by Halton.

The $p$ sets $\left(\left\{\frac{k}{p^2}\right\}\right), \cdots, \left\{\frac{k^s}{p^2}\right\}(1 \leqslant k \leqslant p^2)$ and $\left(\left\{\frac{k}{p}\right\}\right), \cdots, \left\{\frac{k^s}{p}\right\}(1 \leqslant k \leqslant p)$ were proposed by N. M. Korobov [1,7] and the $p$ set $\left(\left\{\frac{k}{p}\right\}\right), \left\{\frac{ak}{p}\right\} \cdots, \left\{\frac{a^{s-1}k}{p}\right\}(1 \leqslant a, k \leqslant p)$ was given by Hua Loo Keny and Wang Yuan [3,7].

Theorems 4.8 and 4.9:Cf. E. Hlawka [4] and N. M. Kolobov [8].

Theorem 4.10: Cf. A. Y. Khintchine [1] and N. S. Bahvalov [1].

Theorem 4.11: Cf. also W. M. schmidt [1] and Wang Yuan [5].

The $\eta$ set of dimension 2 was proposed by N. S. Bahvalov [1] and Hua Loo Keng and Wang Yuan [1,2] independently. (Cf. also S. Haber and C. F. Osgood [1], S. K. Zeremba [1]). Concerning Theorem 4.24, S. K. Zaremba [1] proved a more precise result $D(F_n) = O(F_n^{-1}\ln F_n)$.

The $\mathscr{R}_s$ set and $\eta$ set of dimension> 2 were given by Hua Loo Keng and Wang Yuan [1,4,5,6,7].

Theorem 4.29: Cf. N. M. Korobov [7] and E. Hlawka [4].

The *glp* set $\left(\left\{\frac{a_1k}{p}\right\}\right), \cdots, \left\{\frac{a_sk}{p}\right\}(1 \leqslant k \leqslant p)$ was first introduced by N. M. Korobov [2] and E. Hlawka [3]. Later, Korobov [4] pointed out that the *glp* set may take the form $\left(\left\{\frac{k}{p}\right\}\right), \left\{\frac{ak}{p}\right\} \cdots, \left\{\frac{a^{s-1}k}{p}\right\}(1 \leqslant k \leqslant p)$. H. Niederreiter [3,4] proved that the prime number $p$ in Theorem 4.29 may be replaced by the composite integer $m$ and that for any prime $p$, there exists a primitive root $g$ mod $p$ such that the set $\left(\left\{\frac{k}{p}\right\}, \left\{\frac{gk}{p}\right\} \cdots, \left\{\frac{g^{s-1}k}{p}\right\}\right)(1 \leqslant k \leqslant p)$ has discrepancy $D(p) = O(p^{-1}(\ln p)^s \ln \ln p)$.

# Chapter 5

# Uniform Distribution and Numerical Integration

## 5.1 The function of bounded variation

Let

$$0 = x_0 < x_1 < \cdots < x_l = 1,$$

$(\sigma)$

$$0 = y_0 < y_1 < \cdots < y_m = 1$$

be any division of $G_2$. Let $f(x, y)$ be a function defined on $G_2$ and

$$\begin{aligned} \Delta_{10} f(x_i, y) &= f(x_{i+1}, y) - f(x_i, y), \\ \Delta_{01} f(x, y_j) &= f(x, y_{j+1}) - f(x, y_i), \\ \Delta_{11} f(x_i, y_i) &= f(x_i, y_j) - f(x_{i+1}, y_j) \\ &\quad - f(x_i, y_{j+1}) + f(x_{i+1}, y_{j+1}). \end{aligned}$$

If the variation

$$V_\sigma = \sum_{i=0}^{l-1} \sum_{j=0}^{m-1} |\Delta_{11} f(x_i, y_j)| + \sum_{i=0}^{l-1} |\Delta_{10} f(x_i, 1)| + \sum_{j=0}^{m-1} |\Delta_{01} f(1, y_j)|$$

has an upper bound which is independent of the choice of $(\sigma)$, then $f$ is called a function of bounded variation in the sense of Harday and Krause. The least upper bound of $V_\sigma$ is called the total variation of $f$ and is denoted by $V(f)$. The class of these functions is denoted by $B_2$. Similarly, we may define $B_s (s > 2)$.

If

1. $f(x', y) - f(x, y)$ has the same sign or equals zero,
2. $f(x, y') - f(x, y)$ has the same sign or equals zero and
3. $f(x, y) - f(x', y) - f(x, y') + f(x', y')$ has the same sign or equals zero for any given $x, y, x', y'$, satisfying $0 \leqslant x < x' \leqslant 1$ and $0 \leqslant y < y' \leqslant 1$, then $f$ is called a

generalized monotonic function and the class of these functions is denoted by $M_2$. Similarly, we may define $M_s(s > 2)$.

If $f \in M_2$, then

$$\begin{aligned} V_\sigma &= \left|\sum_{i=0}^{l-1}\sum_{j=0}^{m-1}\Delta_{11}f(x_i,y_j)\right| + \left|\sum_{i=0}^{l-1}\Delta_{10}f(x_i,1)\right| + \left|\sum_{j=0}^{m-1}\Delta_{01}f(1,y_j)\right| \\ &= |f(0,0)-f(0,1)-f(1,0)+f(1,1)| + |f(1,1) \\ &\quad -f(0,1)| + |f(1,1)-f(1,0)|. \end{aligned}$$

Hence $f \in B_2$ and so $M_2 \subset B_2$. Similarly, we may prove that $M_s \subset B_s(s > 2)$.

**Theorem 5.1** *Every function of $B_s$ can be represented as a difference of two functions of $M_s$.*

*Proof.* We prove the theorem only for the case $s = 2$, since the proof is similar for $s > 2$.

For any division $(\sigma)$ of $G_2$, consider a part of $V_\sigma$

$$\sum_{i=1}^{l-1}\sum_{j=0}^{m-1}|\Delta_{11}f(x_i,y_j)|.$$

Change the notations $x_0, y_0$ to $x, y$ respectively, We use $P(x, y)$ and $N(x, y)$ to denote the sums of those terms in above formula satisfying $\Delta_{11}f(x_i, y_j) \geqslant 0$ and $\Delta_{11}f(x_i, y_i) < 0$ respectively. Then

$$\begin{aligned} P(x,y)-N(x,y) &= \sum_{i=0}^{l-1}\sum_{j=0}^{m-1}\Delta_{11}f(x_i,y_j) \\ &= f(1,1)-f(x,1)-f(1,y)+f(x,y), \end{aligned}$$

$$f(x,y) = P(x,y)-N(x,y)+f(x,1)+f(1,y)-f(1,1). \tag{5.1}$$

Since the functions $P(x, y)$ and $N(x, y)$ satisfy

$$\begin{aligned} &\Delta_{10}P \leqslant 0, \quad \Delta_{01}P \leqslant 0, \quad \Delta_{11}P \geqslant 0, \\ &\Delta_{10}N \leqslant 0 \quad \Delta_{01}N \leqslant 0, \quad \Delta_{11}N \geqslant 0, \end{aligned} \tag{5.2}$$

the functions $P$ and $N$ are all generalized monotonic functions.

Since $f(x, 1)$ and $f(1, y)$ are functions of bounded variation of a single variable, they are differences of two monotonic functions, i.e.,

$$\left.\begin{aligned} f(x,1) &= P_1(x,1)-N_1(x,1), \\ f(1,y) &= P_2(1,y)-N_2(1,y), \end{aligned}\right\} \tag{5.3}$$

where $P_1, N_1, P_2, N_2$ are monotonic decreasing functions which may be defined as before.

Substituting (5.3) into (5.1), we have

$$f = F - G,$$

where

$$F = -(N + N_1 + N_2) - f(1,1), \quad G = -(P + P_1 + P_2).$$

From (5.2) and (5.3), we have

$$\begin{aligned} &\Delta_{10}F \geqslant 0, \quad \Delta_{01}F \geqslant 0, \quad \Delta_{11}F \leqslant 0, \\ &\Delta_{10}G \geqslant 0, \quad \Delta_{01}G \geqslant 0, \quad \Delta_{11}G \leqslant 0, \end{aligned}$$

i.e., $F$ and $G$ are generalized monotonic functions. The theorem is proved.

If there exists a positive constant $L$ such that

1. $|f(x',1) - f(x,1)| \leqslant L(x'-x)$,
2. $|f(1,y') - f(1,y)| \leqslant L(y'-y)$,

and

3. $|f(x,y) - f(x',y) - f(x,y') + f(x',y')| \leqslant L(x'-x)(y'-y)$

hold for any given $x, y, x', y'$ satisfying $0 \leqslant x < x' \leqslant 1$ and $0 \leqslant y < y' \leqslant 1$, then $f$ is said to be a function satisfying the generalized Lipschitz condition and the class of these functions is denoted by $L_2$. Similarly, we may define $L_s (s > 2)$.

If $f \in L_2$, then

$$\begin{aligned} V_\sigma &= \Big|\sum_{i=0}^{l-1}\sum_{j=0}^{m-1} \Delta_{11} f(x_i, y_j)\Big| + \Big|\sum_{i=0}^{l-1} \Delta_{10} f(x_i, 1)\Big| + \Big|\sum_{j=0}^{m-1} \Delta_{01} f(1, y_j)\Big| \\ &\leqslant L \sum_{i=0}^{l-1}\sum_{j=0}^{m-1} (x_{i+1} - x_i)(y_{j+1} - y_j) + L\sum_{i=0}^{l-1}(x_{i+1} - x_i) \\ &\quad + L\sum_{j=0}^{m-1}(y_{j+1} - y_j) = 3L \end{aligned}$$

and so $f \in B_2$ and $L_2 \subset B_2$. Similarly, we may prove $L_s \subset B_s (s > 2)$.

Similar to Theorem 5.1, we have

**Theorem 5.2** *Any function of $L_s$ may be represented as a difference of two generalized monotonic functions which satisfy the generalized Lipschitz condition.*

If $f(\mathbf{x})$ satisfies

$$|f| \leqslant L, \quad \left|\frac{\partial f}{\partial x_i}\right| \leqslant L(1 \leqslant i \leqslant s), \quad \left|\frac{\partial^2 f}{\partial x_i \partial x_j}\right| \leqslant L(1 \leqslant i < j \leqslant s), \\ \cdots, \left|\frac{\partial^s f}{\partial x_1 \cdots \partial x_s}\right| \leqslant L, \tag{5.4}$$

then $f \in L_s$ evidently.

## 5.2 Uniform distribution and numerical integration

**Theorem 5.3** *Let $P_n(k)(1 \leqslant k \leqslant n)$ be a set with discrepancy $D(n)$. If $f \in B_s$, then*

$$\left|\int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n}\sum_{k=1}^{n} f(P_n(k))\right| \leqslant V(f)D(n). \tag{5.5}$$

*Proof.* We will prove the theorem only for the case $s = 2$, since the proof is similar for $s > 2$.

Since $f \in B_2$, it follows by Theorem 5.1 that $f$ can be represented as

$$\begin{aligned} &f = F - G, \\ &F = -(N + N_1 + N_2) - f(1,1), \\ &G = -(P + P_1 + P_2), \end{aligned}$$

where

$$\begin{aligned} &\Delta_{10}F \geqslant 0, \Delta_{01}F \geqslant 0, \Delta_{11}F \leqslant 0, \Delta_{10}G \geqslant 0, \Delta_{01}G \geqslant 0, \Delta_{11}G \leqslant 0, \\ &\Delta_{10}P_2 = \Delta_{10}N_2 = \Delta_{01}P_1 = \Delta_{01}N_1 = 0, \\ &\Delta_{11}P_1 = \Delta_{11}N_1 = \Delta_{11}P_2 = \Delta_{11}N_2 = 0, \\ &P(x,1) = N(x,1) = P(1,y) = N(1,y) = 0, \\ &P_1(1,1) = N_1(1,1) = P_2(1,1) = N_2(1,1) = 0. \end{aligned} \tag{5.6}$$

The number of points $P_n(k)(1 \leqslant k \leqslant n)$ which fall in the rectangle

$$\frac{i-1}{q} \leqslant x < \frac{i}{q}, \quad \frac{j-1}{q} \leqslant y < \frac{j}{q} \tag{5.7}$$

is equal to

$$N_n\left(\frac{i}{q}, \frac{j}{q}\right) - N_n\left(\frac{i-1}{q}, \frac{j}{q}\right) - N_n\left(\frac{i}{q}, \frac{j-1}{q}\right) \\ + N_n\left(\frac{i-1}{q}, \frac{j-1}{q}\right).$$

It follows from (5.6) that $F(x,y) \leqslant F\left(\frac{i}{q}, \frac{j}{q}\right)$ for $(x,y)$ belonging to (5.7). Hence

$$S_1 = \frac{1}{n}\sum_{k=1}^{n} F(x_1^{(n)}(k), x_2^{(n)}(k))$$

$$\leqslant \frac{1}{n}\sum_{i=1}^{q}\sum_{j=1}^{q}\Big(N_n\Big(\frac{i}{q},\frac{j}{q}\Big) - N_n\Big(\frac{i-1}{q},\frac{j}{q}\Big)$$
$$- N_n\Big(\frac{i}{q},\frac{j-1}{q}\Big) + N_n\Big(\frac{i-1}{q},\frac{j-1}{q}\Big)\Big)F\Big(\frac{i}{q},\frac{j}{q}\Big)$$
$$=\frac{1}{n}\sum_{i=1}^{q-1}\sum_{j=1}^{q-1} N_n\Big(\frac{i}{q},\frac{j}{q}\Big)\Big(F\Big(\frac{i}{q},\frac{j}{q}\Big) - F\Big(\frac{i+1}{q},\frac{j}{q}\Big)$$
$$- F\Big(\frac{i}{q},\frac{j+1}{q}\Big) + F\Big(\frac{i+1}{q},\frac{j+1}{q}\Big)\Big)$$
$$+ \frac{1}{n}\sum_{i=1}^{q-1} N_n\Big(\frac{i}{q},1\Big)\Big(F\Big(\frac{i}{q},1\Big) - F\Big(\frac{i+1}{q},1\Big)\Big)$$
$$+ \frac{1}{n}\sum_{j=1}^{q-1} N_n\Big(1,\frac{j}{q}\Big)\Big(F\Big(1,\frac{j}{q}\Big) - F\Big(1,\frac{j+1}{q}\Big)\Big)$$
$$+ F(1,1).$$

Since

$$N_n(x,y) = xyn + \psi D(n)n,$$

where we use $\psi$ to denote a number with absolute value $\leqslant 1$, then

$$S_1 \leqslant \sum_{i=1}^{q-1}\sum_{j=1}^{q-1}\Big(\frac{ij}{q^2} + \psi D(n)\Big)\Big(F\Big(\frac{i}{q},\frac{j}{q}\Big) - F\Big(\frac{i+1}{q},\frac{j}{q}\Big)$$
$$- F\Big(\frac{i}{q},\frac{j+1}{q}\Big) + F\Big(\frac{i+1}{q},\frac{j+1}{q}\Big)\Big)$$
$$+ \sum_{i=1}^{q-1}\Big(\frac{i}{q} + \psi D(n)\Big)\Big(F\Big(\frac{i}{q},1\Big) - F\Big(\frac{i+1}{q},1\Big)\Big)$$
$$+ \sum_{j=1}^{q-1}\Big(\frac{j}{q} + \psi D(n)\Big)\Big(F\Big(1,\frac{j}{q}\Big) - F\Big(1,\frac{j+1}{q}\Big)\Big) + F(1,1)$$
$$\leqslant \frac{1}{q^2}\sum_{i=1}^{q}\sum_{j=1}^{q} F\Big(\frac{i}{q},\frac{j}{q}\Big) + V(f)D(n).$$

Let $q \to \infty$. Then

$$S_1 \leqslant \int_0^1\int_0^1 F(x_1,x_2)dx_1dx_2 + V(F)D(n).$$

Similarly, we may prove

$$S_1 \geqslant \int_0^1 \int_0^1 F(x_1, x_2) dx_1 dx_2 - V(F)D(n).$$

Hence

$$S_1 = \int_0^1 \int_0^1 F(x_1, x_2) dx_1 dx_2 + \psi V(F)D(n). \tag{5.8}$$

Similarly,

$$\begin{aligned} S_2 &= \frac{1}{n} \sum_{k=1}^{n} G(x_1^{(n)}(k), x_2^{(n)}(k)) \\ &= \int_0^1 \int_0^1 G(x_1, x_2) dx_1 dx_2 + \psi V(G)D(n). \end{aligned} \tag{5.9}$$

For any given division $(\sigma)$, it follows from (5.6) that

$$\begin{aligned} \sum_{i=0}^{l-1} \sum_{j=0}^{m-1} |\Delta_{11} f(x_i, y_i)| &= P(0,0) - P(0,1) - P(1,0) \\ &\quad + P(1,1) + N(0,0) - N(0,1) - N(1,0) + N(1,1) \\ &= \sum_{i=0}^{l-1} \sum_{j=0}^{m-1} |\Delta_{11} P(x_i, y_j)| + \sum_{i=0}^{l-1} \sum_{j=0}^{m-1} |\Delta_{11} N(x_i, y_j)| \\ &= \sum_{i=0}^{l-1} \sum_{j=0}^{m-1} |\Delta_{11} F(x_i, y_j)| + \sum_{i=0}^{l-1} \sum_{j=0}^{m-1} |\Delta_{11} G(x_i, y_j)|. \end{aligned}$$

Similarly

$$\sum_{i=0}^{l-1} |\Delta_{10} f(x_i, 1)| = \sum_{i=0}^{l-1} |\Delta_{10} F(x_i, 1)| + \sum_{i=0}^{l-1} |\Delta_{10} G(x_i, 1)|$$

and

$$\sum_{j=0}^{m-1} |\Delta_{01} f(1, y_j)| = \sum_{j=0}^{m-1} |\Delta_{01} F(1, y_j)| + \sum_{j=0}^{m-1} |\Delta_{01} G(1, y_j)|.$$

Hence

$$V(f) = V(F) + V(G)$$

and so from (5.8) and (5.9),

$$\left| \int_0^1 \int_0^1 f(x_1, x_2) dx_1 dx_2 - \frac{1}{n} \sum_{k=1}^{n} f(x_1^{(n)}(k), x_2^{(n)}(k)) \right| \leqslant V(f)D(n).$$

The theorem is proved.

**Theorem 5.4** *If* (5.5) *holds for all* $f \in B_s$, *then* $P_n(k) (1 \leqslant k \leqslant n)$ *is a set with discrepancy* $D(n)$.

*Proof.* Let $f(\mathbf{x})$ be the characteristic function of the domain

$$(\mathscr{R}) \qquad 0 \leqslant x_1 < \gamma_1, \cdots, 0 \leqslant x_s < \gamma_s, \quad \gamma \in G_s,$$

i.e.,

$$f(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \mathscr{R}, \\ 0, & \text{if } \mathbf{x} \bar{\in} \mathscr{R}. \end{cases}$$

Then $V(f) \leqslant 1$,

$$\int_{G_s} f(\mathbf{x}) d\mathbf{x} = \int_{\mathscr{R}} f(\mathbf{x}) d\mathbf{x} = |\gamma|$$

and

$$\frac{1}{n} \sum_{k=1}^{n} f(P_n(k)) = \frac{1}{n} N_n(\gamma).$$

Hence it follows from (5.5) that

$$\left| \frac{1}{n} N_n(\gamma) - |\gamma| \right| \leqslant D(n).$$

The theorem is proved.

If the inequality

$$\frac{1}{q^s} \sum_{l_1=1}^{q} \cdots \sum_{l_s=1}^{q} q\delta_{l_1 q} + \cdots + \delta_{l_s q} \left| \frac{1}{n} N_n\left(\frac{l_1}{q}, \cdots, \frac{l_s}{q}\right) - \frac{l_1 \cdots l_s}{q^s} \right| \leqslant D^*(n)$$

holds for any given integer $q \geqslant 1$, then the set of points $P_n(k) (1 \leqslant k \leqslant n)$ is called a set with average discrepancy $D^*(n)$.

**Theorem 5.5** *If $f \in L_s$, then*

$$\left| \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{k=1}^{n} f(P_n(k)) \right| \leqslant L D^*(n).$$

*Proof.* We still suppose that $s = 2$. Since

$$|f(x', y) - f(x, y)| \leqslant |f(x', y) - f(x, y) - f(x', 1) + f(x, 1)| \\ + |f(x', 1) - f(x, 1)| \leqslant 2L|x' - x|$$

and

$$|f(x, y') - f(x, y)| \leqslant 2L|y' - y|,$$

therefore $f$ is uniformly continuous on $G_2$. Hence for any $\varepsilon > 0$,

$$\begin{aligned} S_1 &= \frac{1}{n}\sum_{k=1}^{n} f(x_1^{(n)}(k), x_2^{(n)}(k)) \\ &= \frac{1}{n}\sum_{i=1}^{q}\sum_{j=1}^{q}\Big(N_n\Big(\frac{i}{q},\frac{j}{q}\Big) - N_n\Big(\frac{i-1}{q},\frac{j}{q}\Big) \\ &\quad -N_n\Big(\frac{i}{q},\frac{j-1}{q}\Big) + N_n\Big(\frac{i-1}{q},\frac{j-1}{q}\Big)\Big)\Big(f\Big(\frac{i}{q},\frac{j}{q}\Big)+\delta\Big), \end{aligned}$$

if $q$ is sufficiently large, where $|\delta| \leqslant \varepsilon/2$. Let $S_2$ be that part of $S_1$ which contains those terms not involving $\delta$. Since

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{q}\sum_{j=1}^{q}\Big(N_n\Big(\frac{i}{q},\frac{j}{q}\Big) - N_n\Big(\frac{i-1}{q},\frac{j}{q}\Big) \\ &-N_n\Big(\frac{i}{q},\frac{j-1}{q}\Big) + N_n\Big(\frac{i-1}{q},\frac{j-1}{q}\Big)\Big) = 1, \end{aligned}$$

therefore

$$|S_1 - S_2| < \varepsilon/2. \tag{5.10}$$

By partial summation,

$$\begin{aligned} S_2 =&\frac{1}{n}\sum_{i=1}^{q-1}\sum_{j=1}^{q-1} N_n\Big(\frac{i}{q},\frac{j}{q}\Big)\Big(f\Big(\frac{i}{q},\frac{j}{q}\Big) - f\Big(\frac{i+1}{q},\frac{j}{q}\Big) \\ &- f\Big(\frac{i}{q},\frac{j+1}{q}\Big) + f\Big(\frac{i+1}{q},\frac{j+1}{q}\Big)\Big) \\ &+ \frac{1}{n}\sum_{i=1}^{q-1} N_n\Big(\frac{i}{q},1\Big)\Big(f\Big(\frac{i}{q},1\Big) - f\Big(\frac{i+1}{q},1\Big)\Big) \\ &+ \frac{1}{n}\sum_{j=1}^{q-1} N_n\Big(1,\frac{j}{q}\Big)\Big(f\Big(1,\frac{j}{q}\Big) - f\Big(1,\frac{j+1}{q}\Big)\Big) \\ &+ f(1,1). \end{aligned}$$

Let

$$S_3 = \frac{1}{q^2}\sum_{i=1}^{q}\sum_{j=1}^{q} f\Big(\frac{i}{q},\frac{j}{q}\Big).$$

Then

$$\begin{aligned}|S_2 - S_3| \leqslant & \sum_{i=1}^{q-1}\sum_{j=1}^{q-1}\left|\frac{1}{n}N_n\left(\frac{i}{q},\frac{j}{q}\right) - \frac{ij}{q^2}\right|\left|f\left(\frac{i}{q},\frac{j}{q}\right)\right. \\ & \left. -f\left(\frac{i+1}{q},\frac{j}{q}\right) - f\left(\frac{i}{q},\frac{j+1}{q}\right) + f\left(\frac{i+1}{q},\frac{j+1}{q}\right)\right| \\ & + \sum_{i=1}^{q-1}\left|\frac{1}{n}N_n\left(\frac{i}{q},1\right) - \frac{i}{q}\right|\left|f\left(\frac{i}{q},1\right) - f\left(\frac{i+1}{q},1\right)\right| \\ & + \sum_{j=1}^{q-1}\left|\frac{1}{n}N_n\left(1,\frac{j}{q}\right) - \frac{j}{q}\right|\left|f\left(1,\frac{j}{q}\right) - f\left(1,\frac{j+1}{q}\right)\right| \\ & \leqslant LD^*(n)\end{aligned} \tag{5.11}$$

and

$$\left|\int_0^1\int_0^1 f(x_1,x_2)dx_1dx_2 - S_3\right| < \frac{\varepsilon}{2}, \tag{5.12}$$

if $q$ is sufficiently large. Hence from (5.10), (5.11) and (5.12),

$$\begin{aligned}\left|\int_0^1\int_0^1 f(x_1,x_2)dx_1dx_2 - S_1\right| & \leqslant |S_1 - S_2| + |S_2 - S_3| \\ + \left|S_3 - \int_0^1\int_0^1 f(x_1,x_2), dx_1dx_2\right| & \leqslant LD^*(n) + \varepsilon.\end{aligned}$$

Since $\varepsilon$ is arbitrary, the theorem follows.

It follows from these theorems that we may use the arithmetic mean of the values of the function $f(\mathbf{x})$ over a set $P_n(k)(1 \leqslant k \leqslant n)$

$$\sum_{k=1}^{n} f(P_n(k))$$

to approximate the definite integral

$$\int_{G_s} f(\mathbf{x})d\mathbf{x}.$$

The difference between them is closely related to the discrepancy of $P_n(k)(1 \leqslant k \leqslant n)$, if $f(\mathbf{x})$ satisfies certain conditions. Hence the problem for finding the best quadrature formula is equivalent to the problem for finding the best uniformly distributed sequence of sets. From the view point of numerical analysis, we demand not only the discrepancy of $P_n(k)(1 \leqslant k \leqslant n)$ should the low but also the $P_n(k)$ should be convenient for computation.

## 5.3 The lower estimation for the error term of quadrature formula

It follows from Theorems 5.3 and 3.6 that for any given set $P_n(k)(1 \leqslant k \leqslant n)$, the estimate

$$\Big|\int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n}\sum_{k=1}^{n} f(P_n(k))\Big|$$
$$\leqslant V(f)2^{-2s-4}(s-1)^{-\frac{s-1}{2}}n^{-1}(\log_2 n)^{\frac{s-1}{2}},$$

cannot hold for all $f \in B_s$.

In this section, we shall prove a more precise result for the lower estimation of the error term of quadrature formula.

Define

$$g(\mathbf{x}) = \delta^{-(2s-1)(q+\lambda)}\prod_{k=1}^{s}((x_k - a_k)(a_k + \delta - x_k))^{q+\lambda},$$

where

$$a_k \leqslant x_k < a_k + \delta, \quad 1 \leqslant k \leqslant s \tag{5.13}$$

and where $q$ is an integer, $0 \leqslant \lambda \leqslant 1$ and $\delta > 0$. Outside (5.13), we define

$$g(\mathbf{x}) = 0.$$

Evidently, the function has the following properties:

1) $g(\mathbf{x})$ has $q$-th continuous derivatives everywhere,

2) for $1 \leqslant k \leqslant s,\ i_j \geqslant 0\ (1 \leqslant j \leqslant s)$ and $i_1 + \cdots + i_s = q$, the limit

$$\lim_{y_k \to x_k} \frac{1}{|y_k - x_k|^\lambda}\Big|\frac{\partial^q f(x_1, \cdots, y_k, \cdots, x_s)}{\partial x_1^{i_1}\cdots\partial y_k^{i_k}\cdots\partial x_s^{i_s}} - \frac{\partial^q g(x_1, \cdots, x_k, \cdots, x_s)}{\partial x_1^{i_1}\cdots\partial x_k^{i_k}\cdots\partial x_s^{i_s}}\Big|$$

exists and does not exceed $c(q, \lambda, s)$ and

3) $\displaystyle\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\mathbf{x})d\mathbf{x}$

$$= \delta^{-(2s-1)(q+\lambda)}\prod_{k=1}^{s}\int_{a_k}^{a_k+\delta}((x_k - a_k)(a_k + \delta - x_k))^{q+\lambda}dx_k$$
$$= \delta^{s+q+\lambda}\Big(\int_0^1 [t(1-t)]^{q+\lambda}dt\Big)^s$$
$$= c(q, \lambda, s)\delta^{s+q+\lambda}.$$

Let $n_0 = [(2n)^{1/s}] + 1 (> (2n)^{1/s})$. Divide $G_s$ into $n_0^s (> 2n)$ equal parallelepipeds like (5.13). If the right hand side of (5.13) is 1, the symbol $<$ should be replaced by $\leqslant$. For any given $n$ points $P_n(k)(k = 1, \cdots, n)$ of $G_s$, there exists at least $n$ parallelepipeds not containing these points and then they are denoted by

$$Q_1, \cdots, Q_n.$$

We use $R$ to denote the complement of $\bigcup_{i=1}^{n} Q_i$ with respect to $G_s$. Let

$$f(\mathbf{x}) = \begin{cases} n_0^{(2s-1)(q+\lambda)} \prod_{i=1}^{s} ((x_i - a_i^{(k)})(a_i^{(k)} + n_0^{-1} - x_i))^{q+\lambda}, \text{ if } \mathbf{x} \in Q_k, \\ 0, \text{ if } \mathbf{x} \in R. \end{cases}$$

Since $P_n(k) \bar{\in} \bigcup_{i=1}^{n} Q_i$ for any $k (1 \leqslant k \leqslant n)$, hence

$$f(P_n(k)) = 0, \quad 1 \leqslant k \leqslant n. \tag{5.14}$$

Clearly, we have

$$\int_{G_s} f(\mathbf{x}) d\mathbf{x} \geqslant c(q, \lambda, s) n \cdot n_0^{-(s+q+\lambda)} \geqslant c(q, \lambda, s) n^{-\frac{q+\lambda}{s}} \tag{5.15}$$

by 3). Hence we have

**Theorem 5.6** *For any given $n$ points $P_n(k)(1 \leqslant k \leqslant n)$ of $G_s$, there exists a function $f(\mathbf{x})$ which has the properties* 1) *and* 2) *such that* (5.14) *and* (5.15) *hold.*

Theorem 5.6 means that for any given set of points $P_n(k)(1 \leqslant k \leqslant n)$ and a set of real numbers $\rho_k (1 \leqslant k \leqslant n)$, the difference between the sum

$$\sum_{k=1}^{n} \rho_k f(P_n(k))$$

and the integral

$$\int_{G_s} f(\mathbf{x}) d\mathbf{x}$$

cannot be expected to the less than $c(q, \lambda, s) n^{-\frac{q+\lambda}{s}}$. Especially, if $f$ satisfies (5.4), then the error term of quadrature formula cannot be expected to be better than $O(n^{-1})$.

## 5.4 The quadrature formulas

Suppose in this section that $f \in B_s$. Denote

$$I(f) = \int_{G_s} f(\mathbf{x})d\mathbf{x}.$$

Then by the results of Chapter 4 and Theorem 5.3, we have the following quadrature formulas:

$$\Big|I(f) - \frac{1}{n}\sum_{l_1=0}^{m-1}\cdots\sum_{l_s=0}^{m-1} f\Big(\frac{l_1}{m},\cdots,\frac{l_s}{m}\Big)\Big| \leqslant V(f)2^s n^{-\frac{1}{s}}, \quad n = m^s, \qquad \text{(Cf. §4.1)}$$

$$\Big|I(f) - \frac{1}{n}\sum_{k=1}^{n} f(\varphi_{p_1}(k),\cdots,\varphi_{p_s}(k))\Big| \leqslant V(f)\Big(\prod_{i=1}^{s}\frac{p_i \ln p_i n}{\ln p_i}\Big)n^{-1}, \qquad \text{(Cf. §4.2)}$$

$$\Big|I(f) - \frac{1}{n}\sum_{k=1}^{n} f\Big(\frac{k}{n},\varphi_{p_1}(k),\cdots,\varphi_{p_{s-1}}(k)\Big)\Big| \leqslant V(f)\Big(\prod_{i=1}^{s-1}\frac{p_i \ln p_i n}{\ln p_i}\Big)n^{-1}, \qquad \text{(Cf. §4.2)}$$

$$\Big|I(f) - \frac{1}{n}\sum_{a=1}^{p}\sum_{k=1}^{p} f\Big(\Big\{\frac{k}{p}\Big\},\Big\{\frac{ak}{p}\Big\},\cdots,\Big\{\frac{a^{s-1}k}{p}\Big\}\Big)\Big| \leqslant V(f)c(s)n^{-\frac{1}{2}}(\ln n)^s, \quad n = p^2. \qquad \text{(Cf. §4.3)}$$

$$\Big|I(f) - \frac{1}{n}\sum_{k=1}^{n} f\Big(\Big\{\frac{k}{n}\Big\},\Big\{\frac{k^2}{n}\Big\},\cdots,\Big\{\frac{k^s}{n}\Big\}\Big)\Big| \leqslant V(f)c(s)n^{-\frac{1}{2}}(\ln n)^s, \quad n = p^2. \qquad \text{(Cf. §4.3)}$$

$$\Big|I(f) - \frac{1}{p}\sum_{k=1}^{p} f\Big(\Big\{\frac{k}{p}\Big\},\Big\{\frac{k^2}{p}\Big\},\cdots,\Big\{\frac{k^s}{p}\Big\}\Big)\Big| \leqslant V(f)p^{-\frac{1}{2}}(\ln p)^s. \qquad \text{(Cf. §4.3)}$$

$$\Big|I(f) - \frac{1}{n}\sum_{k=1}^{n} f(\{\alpha_1 k\},\cdots,\{\alpha_s k\})\Big| \leqslant V(f)c(\boldsymbol{\alpha},\varepsilon)n^{-1+\varepsilon}. \qquad \text{(Cf. §4.5)}$$

$$\Big|I(f) - \frac{1}{n}\sum_{k=1}^{n} f(\{\beta_1 k\},\cdots,\{\beta_s k\})\Big| \leqslant V(f)c(\boldsymbol{\beta},\varepsilon)n^{-1+\varepsilon}. \qquad \text{(Cf. §4.5)}$$

$$\begin{aligned}&\Big|I(f)-\frac{1}{n}\sum_{k=1}^{n}f\Big(\Big\{\frac{c_1k}{n}\Big\},\cdots,\Big\{\frac{c_sk}{n}\Big\}\Big)\Big|\\&\leqslant V(f)c(\mathscr{R}_s,\varepsilon)n^{-\frac{1}{2}-\frac{1}{2(s-1)}+\varepsilon},\\&s=\frac{\varphi(m)}{2},\quad n=n_l,\quad c_i=c_{li}\\&(l=1,2,\cdots,\quad i=1,\cdots,s).\end{aligned} \qquad \text{(Cf. §4.6)}$$

$$\begin{aligned}&\Big|I(f)-\frac{1}{q}\sum_{k=1}^{q}f\Big(\Big\{\frac{c_2k}{n}\Big\},\cdots,\Big\{\frac{c_{s+1}k}{n}\Big\}\Big)\Big|\\&\leqslant V(f)c(\mathscr{R}_{s+1},\varepsilon)q^{-1+\varepsilon},\\&s=\frac{\varphi(m)}{2}-1,\quad q=[n^{\frac{1}{2}+\frac{1}{2s}}].\end{aligned} \qquad \text{(Cf. §4.6)}$$

$$\begin{aligned}&\Big|I(f)-\frac{1}{F_n}\sum_{k=1}^{F_n}f\Big(\Big\{\frac{k}{F_n}\Big\},\Big\{\frac{F_n(2)}{F_n}k\Big\},\cdots,\Big\{\frac{F_n(s)}{F_n}k\Big\}\Big)\Big|\\&\leqslant V(f)c(\eta)F_n^{-\frac{1}{2}-\frac{1}{2^{s+1}\ln 2}-\frac{1}{2^{2s+3}}},\quad F_n=F_{s,n}.\end{aligned} \qquad \text{(Cf. §4.7)}$$

$$\begin{aligned}&\Big|I(f)-\frac{1}{F_n}\sum_{k=1}^{F_n}f\Big(\Big\{\frac{k}{F_n}\Big\},\Big\{\frac{F_{n-1}}{F_n}k\Big\}\Big)\Big|\\&\leqslant V(f)cF_n^{-1}(\ln F_n)^2,\quad F_n=F_{2,n}.\end{aligned} \qquad \text{(Cf. §4.7)}$$

$$\begin{aligned}&\Big|I(f)-\frac{1}{q}\sum_{k=1}^{q}f\Big(\Big\{\frac{F_n(2)}{F_n}k\Big\},\cdots,\Big\{\frac{F_n(s+1)}{F_n}k\Big\}\Big)\Big|\\&\leqslant V(f)c(\eta,\varepsilon)q^{-1+\varepsilon},\quad F_n=F_{s+1,n},\\&q=[F_n^{\frac{1}{2}+\frac{1}{2^{s+2}\ln 2}+\frac{1}{2^{2s+4}}}].\end{aligned} \qquad \text{(Cf. §4.7)}$$

$$\begin{aligned}&\Big|I(f)-\frac{1}{p}\sum_{k=1}^{p}f\Big(\Big\{\frac{a_1k}{p}\Big\},\cdots,\Big\{\frac{a_sk}{p}\Big\}\Big)\Big|\\&\leqslant V(f)c(s)p^{-1}(\ln p)^s.\end{aligned} \qquad \text{(Cf. §4.9)}$$

$$\begin{aligned}&\Big|I(f)-\frac{1}{p}\sum_{k=1}^{n}f\Big(\Big\{\frac{a_2k}{p}\Big\},\cdots,\Big\{\frac{a_{s+1}k}{p}\Big\}\Big)\Big|\\&\leqslant V(f)c(s)n^{-1}(\ln p)^{s+1},\quad 1\leqslant n\leqslant p.\end{aligned} \qquad \text{(Cf. §4.9)}$$

## Notes

The definition for a function of bounded variation was given by M. Krause [1] and G. H. Hardy [1] (Cf. also C. R. Adams, and J. A. Clarkson [1,2] and S. K. Zaremba [2]).

Theorem 5.3 was proved by J. F. Koksma [1] for $s = 1$ and generalized to $s > 1$ by E. Hlawka [1] (Cf. also E. M. Sobol [1] for the class of functions $L_s$).

Theorem 5.6: Cf. N. S. Bahvalov [1].

# Chapter 6

# Periodic Functions

## 6.1 The classes of functions

The $G_s$ may be regarded as tori. The 1-dimensional torus $G_1$ may be obtained by identifying two end-points of the unit interval $0 \leqslant x_1 \leqslant 1$ and $G_2$ by identifying 2 opposite sides of the unit square $0 \leqslant x_1 \leqslant 1$, $0 \leqslant x_2 \leqslant 1$. In general, $G_s$ is obtained by identifying the $2s$ opposite surfaces of the $s$-dimensional unit cube, i.e., the points

$$(x_1, \cdots, x_{v-1}, 0, x_{v+1}, \cdots x_s)$$

and

$$(x_1, \cdots, x_{v-1}, 1, x_{v+1}, \cdots, x_s)$$

are identified, where $1 \leqslant v \leqslant s$.

Hereafter, we shall use $G_s$ to denote the $s$-dimensional torus, if not otherwise specified.

If a single-valued function $f(x_1, \cdots, x_s)$ is a periodic function of $s$ variables each with period 1, i.e.,

$$f(x_1, \cdots, x_v + 1, \cdots, x_s) = f(x_1, \cdots, x_v, \cdots, x_s), \quad 1 \leqslant v \leqslant s,$$

then we have a single-valued function over $G_s$. A simple example is $e^{2\pi i(m_1x_1+\cdots+m_sx_s)}$, where $m_1, \cdots, m_s$ are integers.

Let $f(x_1, \cdots, x_s)$ be a single-valued function of $G_s$. Let $\alpha = \rho + \beta$, where $\rho$ is a non-negative integer and $0 < \beta \leqslant 1$. Put

$$\begin{aligned}\delta_{h,k} f =& (2i)^{-1}(f(x_1, \cdots, x_k + h, \cdots, x_s) \\ &-f(x_1, \cdots, x_k - h, \cdots, x_s)).\end{aligned}$$

Suppose that the derivatives

$$\frac{\partial^{\tau_1+\cdots+\tau_s} f}{\partial x_1^{\tau_1} \cdots \partial x_s^{\tau_s}} = f^{(\tau_1\cdots,\tau_s)}, \quad 0 \leqslant \tau_1, \cdots, \tau_s \leqslant \rho$$

exist and are the periodic functions of $s$ variables each with period 1. Let

$$||f^{\alpha}|| = \sup_{\substack{0<h_k\leqslant 1\\ \mathbf{x}\in G_s}} \left|\left(\left(\prod_{k=1}^{s} h_k^{-\beta}\delta_{h_k,k}\right)f\right)^{(\rho,\cdots,\rho)}\right|.$$

Let $H_s^{\alpha}(C)$ denote the class of functions $f(\mathbf{x})$ of $G_s$, such that

$$||f^{\alpha}|| \leqslant C,$$

where as usual, $C$ denotes the absolute constant and the lower derivatives of $f(\mathbf{x})$ are also bounded by $C$. Let

$$\mu(x) = \begin{cases} \left(\cos\left(\frac{\pi}{2}\log_2|x|\right)\right)^2, \text{if } \frac{1}{2} \leqslant |x| \leqslant 2, \\ 0, \text{ otherwise,} \end{cases}$$
$$\mu(x) = \mu(2^{1-t}x)$$

for any given positive integer $t$ and

$$\mu_0(x) = 1 - \sum_{t=1}^{\infty} \mu_t(x). \tag{6.1}$$

Then

$$\mu(x) + \mu\left(\frac{x}{2}\right) = \left(\cos\left(\frac{\pi}{2}\log_2|x|\right)\right)^2 + \left(\sin\left(\frac{\pi}{2}\log_2|x|\right)\right)^2 = 1$$

for $1 \leqslant |x| \leqslant 2$ and so

$$\sum_{t=1}^{\infty} \mu_t(x) = 1$$

or

$$\mu_0(x) = 0$$

for$|x| \geqslant 1$.

Suppose that $f(\mathbf{x})$ has the Fourier expansion

$$f(\mathbf{x}) \sim \sum C(\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})},$$

where $\mathbf{m}$ runs over all the integral vectors. If the series

$$\sum C(\mathbf{m})\lambda(\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})}$$

converges everywhere, then its sum is denoted by $f(\mathbf{x}) \odot \lambda(\mathbf{m})$. For any given non-negative integral vector $\mathbf{t} = (t_1, \cdots, t_s)$(i.e.,$t_i \leqslant 0, 1 \leqslant i \leqslant s$), Let

$$t_0 = t_1 + \cdots + t_s$$

and

$$\varphi_{\mathbf{t}}(\mathbf{x}) = f(\mathbf{x}) \bigodot \prod_{k=1}^{s} \mu_{t_k}(m_k) = \sum C_{\mathbf{t}}(\mathbf{m}) e^{2\pi i(\mathbf{m},\mathbf{x})},$$

where

$$C_{\mathbf{t}}(\mathbf{m}) = C(\mathbf{m}) \mu_{t_1}(m_1) \cdots \mu_{t_s}(m_s).$$

Let $Q_s^\alpha(C)$ denote the class of continuous function $f(\mathbf{x})$ of $G_s$ satisfying

$$||\varphi_t|| = \sup_{\mathbf{x} \in G_s} |\varphi_{\mathbf{t}}| \leqslant C 2^{-\alpha t_0}.$$

Let $E_s^\alpha(C)$ denote the set of functions $f(\mathbf{x})$ of $G_s$ such that

$$f(\mathbf{x}) \sim \sum C(\mathbf{m}) e^{2\pi i(\mathbf{m},\mathbf{x})},$$

where

$$|C(\mathbf{m})| \leqslant \frac{C}{||\mathbf{m}||^\alpha}.$$

## 6.2 Several lemmas

**Lemma 6.1** *Let*

$$\Delta_2 \lambda(n) = \lambda(n+1) - 2\lambda(n) + \lambda(n-1) \tag{6.2}$$

*and*

$$K_\lambda(x) = \sum \lambda(n) e^{2\pi i n x}.$$

*If $\lambda(n) = 0$ for $|n| \leqslant M$, where $M$ is a positive integer, then*

$$\sum \Delta_2 \lambda(n) = 0, \tag{6.3}$$

$$\sum n \Delta_2 \lambda(n) = 0 \tag{6.4}$$

*and*

$$K_\lambda(x) = -\sum \Delta_2 \lambda(n) \frac{e^{2\pi i n x}}{4(\sin \pi x)^2}. \tag{6.5}$$

*Proof.* (6.3) and (6.4) immediately follow from (6.2). Now we proceed to prove (6.5):

$$\begin{aligned} -\sum \Delta_2 \lambda(n) \frac{e^{2\pi i n x}}{4(\sin \pi x)^2} &= -\sum \lambda(n) e^{2\pi i n x} \frac{(e^{2\pi i x} - 2 + e^{-2\pi i x})}{4(\sin \pi x)^2} \\ &= -\sum \lambda(n) e^{2\pi i n x} \frac{(e^{\pi i x} - e^{-\pi i x})^2}{4(\sin \pi x)^2} = K_\lambda(x). \end{aligned}$$

The lemma is proved.

Put

$$||f(\mathbf{x})||_1 = \int_{G_s} |f(\mathbf{x})| d\mathbf{x}.$$

**Lemma 6.2** *Under the assumption of Lemma* 6.1, *we have*

$$||K_\lambda(x)||_1 \leqslant \frac{\pi M}{2} \sum |\Delta_2 \lambda(n)|.$$

*Proof.* Let

$$||K_\lambda(x)||_1 = I_1 + I_2, \tag{6.6}$$

where

$$I_1 = \int_\varepsilon^{1-\varepsilon} |K_\lambda(x)| dx, \quad I_2 = \int_{-\varepsilon}^{\varepsilon} |K_\lambda(x)| dx,$$

in which $\varepsilon = \frac{1}{2\pi M}$. Since $\sin \pi x \geqslant 2x$ for $0 \leqslant x \leqslant \frac{1}{2}$, therefore

$$\int_\varepsilon^{1-\varepsilon} \left| \frac{e^{2\pi i n x}}{4(\sin \pi x)^2} \right| dx \leqslant 2 \int_\varepsilon^{1/2} \frac{dx}{4(\sin \pi x)^2} \leqslant \frac{1}{8} \int_\varepsilon^{1/2} \frac{dx}{x^2} < \frac{1}{8\varepsilon} = \frac{\pi M}{4}.$$

Hence

$$I_1 \leqslant \frac{\pi M}{4} \sum |\Delta_2 \lambda(n)| \tag{6.7}$$

by (6.5). Since

$$K_\lambda(x) = -\sum \Delta_2 \lambda(n) \frac{(e^{2\pi i n x} - 1 - 2\pi i n x)}{4(\sin \pi x)^2}$$

by Lemma 6.1 and

$$|e^{i\alpha} - 1 - i\alpha| \leqslant \alpha^2$$

for $-1 \leqslant \alpha \leqslant 1$, therefore

$$\int_{-\varepsilon}^{\varepsilon} \left| \frac{e^{2\pi i n x} - 1 - 2\pi i n x}{4(\sin \pi x)^2} \right| dx \leqslant \pi^2 n^2 \int_{-\varepsilon}^{\varepsilon} \left( \frac{x}{\sin \pi x} \right)^2 dx$$
$$\leqslant \frac{\pi^2 n^2 \varepsilon}{2} = \frac{\pi M}{4}$$

for $|n| \leqslant M$. Hence

$$I_2 \leqslant \frac{\pi M}{4} \sum |\Delta_2 \lambda(n)|. \tag{6.8}$$

The lemma follows from (6.6), (6.7) and (6.8).

**Lemma 6.3** *Suppose that* $f(\mathbf{x})$ *has the Fourier expansion*

$$f(\mathbf{x}) \sim \sum C(\mathbf{m}) e^{2\pi i (\mathbf{m}, \mathbf{x})}.$$

*Then*

$$||f(\mathbf{x}) \bigodot \lambda(\mathbf{m})|| \leqslant ||f(\mathbf{x})|| \prod_{k=1}^{s} ||K_{\lambda k}(x)||_1,$$

*where*

$$\lambda(\mathbf{m}) = \lambda_1(m_1)\cdots\lambda_s(m_s).$$

*Proof.* Since

$$\int_0^1 e^{2\pi inx}dx = \begin{cases} 1, & \text{if } n = 0, \\ 0, & \text{if } n \neq 0, \end{cases} \tag{6.9}$$

so

$$\begin{aligned} f(\mathbf{x}) \bigodot \lambda(\mathbf{m}) &= \sum C(\mathbf{m})\lambda(\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})} \\ &= \int_{G_s} \prod_{v=1}^{s} K_{\lambda_v}(x_v - z_v)f(\mathbf{z})d\mathbf{z} \\ &= (-1)^s \int_{G_s} \prod_{v=1}^{s} K_{\lambda_v}(y_v)f(\mathbf{x}-\mathbf{y})d\mathbf{y} \end{aligned}$$

and

$$||f(\mathbf{x}) \bigodot \lambda(\mathbf{m})|| \leqslant ||f(\mathbf{x})|| \prod_{k=1}^{s} ||K_{\lambda_k}(x)||_1.$$

The lemma is proved.

**lemma 6.4**

$$d(\mu) = \int_{-\infty}^{\infty} (|\mu| + 2|\mu'|)dx + V(\mu') < c,$$

*where c is a positive constant and V(f) denotes the total variation of f(x) in the interval* $(-\infty, \infty)$.

*Proof.* Since

$$\mu'(x) = \begin{cases} -\dfrac{\pi}{|x|\ln 2}\cos\left(\dfrac{\pi}{2}\log_2|x|\right)\sin\left(\dfrac{\pi}{2}\log_2|x|\right), \\ \quad \text{if } \dfrac{1}{2} \leqslant x \leqslant 2, \\ 0, \text{ otherwise,} \end{cases}$$

$\mu'(x)$ is therefore a product of the functions of bounded variation and so $\mu'(x)$ is a function of bounded variation too. The lemma is proved.

**Lemma 6.5** *Let*

$$g(\psi) = \max_{\frac{1}{2} \leqslant |x| \leqslant 2} (|\psi|, |\psi'|, |\psi''|).$$

*Then*

$$||K_{\psi(n2^{1-t})\mu(n2^{1-t})}(x)||_1 \leqslant 2\pi d(\mu)g(\psi).$$

*Proof.* Since

$$\begin{aligned}&|\Delta_2(\psi(n2^{1-t})\mu(n2^{1-t}))|\\&=\Big|\int_0^{2^{1-t}}((\psi\mu)'(n2^{1-t}+z)-(\psi\mu)'((n-1)2^{1-t}+z))dz\Big|\\&\leqslant 2^{1-t}\max_{0\leqslant z\leqslant 2^{1-t}}|(\psi\mu)'(n2^{1-t}+z)-(\psi\mu)'((n-1)2^{1-t}+z)|\end{aligned}$$

and $\mu(n2^{1-t})=0$ for $|n|\geqslant 2^t$, therefore by Lemma 6.2,

$$\begin{aligned}&||K_{\psi(n2^{1-t})\mu(n2^{1-t})}(x)||_1\leqslant \pi 2^{t-1}\sum|\Delta_2(\psi(n2^{1-t})\mu(n2^{1-t}))|\\&\leqslant \pi\sum\max_{0\leqslant z\leqslant 2^{1-t}}|(\psi\mu)'(n2^{1-t}+z)-(\psi\mu)'((n-1)2^{1-t}+z)| \qquad (6.10)\\&\leqslant 2\pi V((\psi\mu)').\end{aligned}$$

Since

$$\begin{aligned}(\psi\mu)'(y)-(\psi\mu)'(x)&=\int_x^y(\psi\mu)''dz\\&=\int_x^y(\psi\mu''+2\psi'\mu'+\psi''\mu)dz,\end{aligned}$$

so

$$V((\psi\mu)')\leqslant g(\psi)d(\mu) \tag{6.11}$$

by Lemma 6.4. The lemma follows by substituting (6.11) into (6.10).

## 6.3 The relations between $H_s^\alpha(C), Q_s^\alpha(C)$ and $E_s^\alpha(C)$

**Theorem 6.1** $H_s^\alpha(C)\subset Q_s^\alpha(C\cdot c(\alpha)^s)$.

*Proof.* Let

$$\mu_{l,t_k}(n)=\mu_{t_k}(n)((2\pi in)^l\sin 2\pi nh_k)^{-1}$$

and

$$K_{l,t_k}(x)=\sum\mu_{l,t_k}(n)e^{2\pi inx},$$

where $h_k=\dfrac{1}{5}2^{-t_k}$. Let $f\in H_s^\alpha(C)$ and

$$f(\mathbf{x})\sim\sum C(\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})}.$$

Since

$$\begin{aligned}\Big(\prod_{k=1}^s\delta_{h_k,k}\Big)f&=\sum C(\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})}\prod_{k=1}^s(2i)^{-1}(e^{2\pi im_kh_k}-e^{-2\pi im_kh_k})\\&=f\odot\prod_{k=1}^s\sin 2\pi m_kh_k,\end{aligned}$$

we have

$$\begin{aligned}\varphi_t &= f \odot \prod_{k=1}^{s} \mu_{t_k}(m_k) = \Big(\prod_{k=1}^{s} \delta_{h_k,k}\Big) f \odot \prod_{k=1}^{s} \mu_{0,t_k}(m_k) \\ &= \int_{G_s} \prod_{k=1}^{s} K_{0,t_k}(x_k - z_k)\Big(\Big(\prod_{j=1}^{s} \delta_{h_j,j}\Big) f(\mathbf{z})\Big) d\mathbf{z}\end{aligned}$$

by (6.9). Since

$$K_{l,t_k}^{(l)}(x) = K_{0,t_k}(x),$$

hence by partial integration $\rho$ times with respect to every variable $z_k$, we have

$$\begin{aligned}\varphi_{\mathbf{t}} &= \int_{G_s} \Big(\prod_{k=1}^{s} K_{\rho,t_k}(x_k - z_k)\Big)\Big(\Big(\prod_{j=1}^{s} \delta_{h_j,j}\Big) f(\mathbf{z})\Big)^{(\rho,\cdots,\rho)} d\mathbf{z} \\ &= \Big(\Big(\prod_{k=1}^{s} \delta_{h_k,k}\Big) f\Big)^{(\rho,\cdots,\rho)} \odot \prod_{j=1}^{s} \mu_{\rho,t_j}(m_j).\end{aligned} \tag{6.12}$$

Since $f \in H_s^\alpha(C)$, therefore

$$\Big\|\Big(\Big(\prod_{k=1}^{s} \delta_{h_k,k}\Big) f\Big)^{(\rho,\cdots,\rho)}\Big\| \leqslant C \prod_{k=1}^{s} h_k^\beta \leqslant C 5^{-\beta s} 2^{-\beta t_0}. \tag{6.13}$$

Put

$$\psi(x) = ((2\pi i x 2^{t_k - 1})^\rho \sin(2\pi x 2^{t_k-1} h_k))^{-1}.$$

Since

$$|\sin(2\pi x 2^{t_k-1} h_k)| \geqslant \sin\frac{\pi}{10}$$

for $\frac{1}{2} \leqslant |x| \leqslant 2$, therefore

$$g(\psi) = \max_{\frac{1}{2} \leqslant |x| \leqslant 2} (|\psi|, |\psi'|, |\psi''|) \leqslant c(\alpha) 2^{-\rho t_k}$$

and so

$$||K_{\rho,t_k}(x)||_1 \leqslant c(\alpha) 2^{-\rho t_k} \tag{6.14}$$

by Lemma 6.5. Hence it follows by (6.12), (6.13), (6.14) and Lemma 6.3 that

$$\begin{aligned}||\varphi_{\mathbf{t}}|| &\leqslant \Big\|\Big(\Big(\prod_{k=1}^{s} \delta_{h_k,k}\Big) f\Big)^{(\rho,\cdots,\rho)}\Big\| \prod_{j=1}^{s} ||K_{\rho,t_j}(x)||_1 \\ &\leqslant Cc(\alpha)^s 2^{-\beta t_0 - \rho t_0} = Cc(\alpha)^s 2^{-\alpha t_0}\end{aligned}$$

and so $f \in Q_s^\alpha(Cc(\alpha)^s)$. The theorem is proved.

**Theorem 6.2**

$$Q_s^\alpha(C) \subset E_s^\alpha(C2^s).$$

*Proof* Suppose that $f \in Q_s^\alpha(C)$ and

$$\varphi_{\mathbf{t}} = f \bigodot \prod_{k=1}^{s} \mu_{t_k}(m_k) = \sum C_{\mathbf{t}}(\mathbf{m}) e^{2\pi i(\mathbf{m},\mathbf{x})}.$$

Since

$$C_{\mathbf{t}}(\mathbf{m}) = \int_{G_s} \varphi_{\mathbf{t}}(x) e^{-2\pi i(\mathbf{m},\mathbf{x})} d\mathbf{x},$$

hence

$$|C_{\mathbf{t}}(\mathbf{m})| \leqslant \int_{G_s} |\varphi_{\mathbf{t}}(\mathbf{x})| d\mathbf{x} \leqslant ||\varphi_{\mathbf{t}}|| \leqslant C2^{-\alpha t_0}.$$

Since $C_{\mathbf{t}}(\mathbf{m})=0$ for $|m_k| \geqslant 2^{t_k}$, therefore

$$|C_{\mathbf{t}}(\mathbf{m})| \leqslant C||\mathbf{m}||^{-\alpha}. \tag{6.15}$$

From (6.1),

$$C(\mathbf{m}) = C(\mathbf{m}) \sum_{t_1=0}^{\infty} \mu_{t_1}(m_1) \cdots \sum_{t_s=0}^{\infty} \mu_{t_s}(m_s) = \sum\nolimits'' C_{\mathbf{t}}(\mathbf{m}), \tag{6.16}$$

where $\sum''$ denotes a sum of which $\mathbf{t}$ runs over all non-negative integral vectors. For any given $\mathbf{m}$, there are at most $2^s$ non-negative integral vectors $\mathbf{t}$ such that

$$2^{-1} < |2^{1-t_k} m_k| < 2, \quad k = 1, \cdots, s,$$

i.e.; $|C_{\mathbf{t}}(\mathbf{m})| \neq 0$. Hence

$$|C(\mathbf{m})| \leqslant C2^s ||\mathbf{m}||^{-\alpha}$$

by (6.15) and (6.16) and so $f \in E_s^\alpha(C2^s)$. The theorem is proved.

**Theorem 6.3** *Suppose that $f \in Q_s^\alpha(C)$. Then*

$$f(\mathbf{x}) = \sum\nolimits'' \varphi_{\mathbf{t}}(\mathbf{x}).$$

*Proof.* Since $f \in Q_s^\alpha(C)$, therefore

$$||\varphi_{\mathbf{t}}|| = \sup_{\mathbf{x} \in G_s} |\varphi_{\mathbf{t}}(\mathbf{x})| \leqslant C2^{-\alpha t_0}$$

and

$$\sum{}'' \|\varphi_{\mathbf{t}}\| \leqslant C\sum{}'' 2^{-\alpha t_0} = C\Big(\sum_{t=0}^{\infty} 2^{-\alpha t}\Big)^s = C_c(\alpha)^s.$$

Hence the series

$$\sum{}'' \varphi_{\mathbf{t}}(\mathbf{x})$$

is uniformly convergent on $G_s$ and so it represents a continuous function on $G_s$ and it is denoted by $f_0(\mathbf{x})$. For any given integral vector $\mathbf{n}$,

$$\begin{aligned}
&\int_{G_s} (f(\mathbf{x}) - f_0(\mathbf{x}))e^{-2\pi i(\mathbf{n},\mathbf{x})}d\mathbf{x} \\
&\quad = \sum \int_{G_s} (C(\mathbf{m}) - \sum{}'' C_{\mathbf{t}}(\mathbf{m}))e^{2\pi i(\mathbf{m}-\mathbf{n},\mathbf{x})}d\mathbf{x} \\
&\quad = C(\mathbf{n}) - \sum{}'' C_{\mathbf{t}}(\mathbf{n}) \\
&\quad = C(\mathbf{n})(1 - \sum{}'' \mu_{t_s}(n_1)\cdots\mu_{t_s}(n_s)) = 0.
\end{aligned}$$

Hence $f(\mathbf{x})$ and $f_0(\mathbf{x})$ are equal almost everwhere on $G_s$. Since they are both continuous functions on $G_s$, we have $f(\mathbf{x}) = f(\mathbf{x}_0)$. The theorem is proved.

## 6.4 Periodic functions

In these next two sections, we shall introduce methods of reducing the integral of a certain class of functions to the integral of a class of periodic functions.

Let $\mathbf{x}_v(x) = (x_1, \cdots, x_{v-1}, x, x_{v+1}, \cdots, x_s)$. Let $f(\mathbf{x})$ be a function on $s$-dimensional unit cube $G_s$. Let $\alpha > 0$ and $\alpha = \rho + \beta$, where $\rho$ is a non-negative integer and $0 < \beta \leqslant 1$. Define

$$\sigma_{h,k} f = f(\mathbf{x}_k(x_k + h)) - f(\mathbf{x})$$

for $\mathbf{x} \in G_s$ and $\mathbf{x}_k(x_k + h) \in G_s$. Suppose that the derivatives

$$\frac{\partial^{\tau_1+\cdots+\tau_s}}{\partial x_1^{\tau_1}\cdots\partial x_s^{\tau_s}} = f^{(\tau_1,\cdots,\tau_s)}, \quad 0 \leqslant \tau_1, \cdots, \tau_s \leqslant \rho$$

exist. Let

$$\|f^\alpha\| = \sup_{\substack{\mathbf{x}\in G_s \\ \mathbf{x}+\mathbf{h}\in G_s}} \Big|\Big(\Big(\prod_{k=1}^{s} |h_k|^{-\beta}\sigma_{h_k,k}\Big)f\Big)^{(\rho,\cdots,\rho)}\Big|$$

and $D_s^\alpha(C)$ be the class of functions of $G_s$ such that

$$\|f^\alpha\| \leqslant C$$

and the lower derivatives of $f$ are also bounded by $C$. For any $f \in D_s^\alpha(C)$, if there exists $\varphi(\mathbf{x}) \in D_s^\alpha(Cc(\alpha)^s)$ such that

$$\varphi(\mathbf{x}_v(1)) = \varphi(\mathbf{x}_v(0)), \quad v = 1, \cdots, s \tag{6.17}$$

and

$$\int_{G_s} f(\mathbf{x})d\mathbf{x} = \int_{G_s} \varphi(\mathbf{x})d\mathbf{x}, \tag{6.18}$$

the $\varphi(\mathbf{x})$ is called a simple periodic function of $f(\mathbf{x})$.

From the definition of $H_s^\alpha(C)$, we derive immediately

**Theorem 6.4** *Suppose that $\alpha \leqslant 1$. If $\varphi \in D_s^\alpha(C)$ is a simple periodic function of $f$, then $\varphi(\{x_1\}, \cdots, \{x_x\}) \in H_s^\alpha(C)$.*

We know from Theorem 6.4 that if $\alpha \leqslant 1$, then the quadrature formula of the functions of $D_s^\alpha(C)$ may be deduced from the quadrature formula of the functions of $H_s^\alpha(C)$. Now we shall introduce several methods for constructing the simple periodic functions.

1. Let

$$\begin{aligned}\varphi_1(\mathbf{x}) &= \frac{1}{2}(f(\mathbf{x}) + f(\mathbf{x}_1(1 - x_1))),\\ \varphi_2(\mathbf{x}) &= \frac{1}{2}(\varphi_1(\mathbf{x}) + \varphi_1(\mathbf{x}_2(1 - x_2))),\\ &\cdots\cdots\\ \varphi_s(\mathbf{x}) &= \frac{1}{2}(\varphi_{s-1}(\mathbf{x}) + \varphi_{s-1}(\mathbf{x}_s(1 - x_s)))\end{aligned} \tag{6.19}$$

and

$$\varphi(\mathbf{x}) = \varphi_s(\mathbf{x}).$$

Then $f \in D_s^\alpha(C)$ obviously implies that $\varphi \in D_s^\alpha(C)$. From (6.19),

$$\varphi_1(\mathbf{x}_1(1)) = \varphi_1(\mathbf{x}_1(0)) = \frac{f(\mathbf{x}_1(1)) + f(\mathbf{x}_1(0))}{2}.$$

Hence by the substitution $1 - x_1 = y_1$, we have

$$\int_{G_s} f(\mathbf{x})d\mathbf{x} = \int_{G_s} \varphi_1(\mathbf{x})d\mathbf{x}.$$

It is easily verified by induction that $\varphi$ satisfies (6.17) and (6.18). Hence $\varphi$ is a simple periodic function of $f$.

2. Let

$$\begin{aligned}\varphi_1(\mathbf{x}) &= f(\mathbf{x}) + \Big(x_1 - \frac{1}{2}\Big)(f(\mathbf{x}_1(0)) - f(\mathbf{x}_1(1))),\\ \varphi_2(\mathbf{x}) &= \varphi_1(\mathbf{x}) + \Big(x_2 - \frac{1}{2}\Big)(\varphi_1(\mathbf{x}_2(0)) - \varphi_1(\mathbf{x}_2(1))),\\ &\cdots\cdots\\ \varphi_s(\mathbf{x}) &= \varphi_{s-1}(\mathbf{x}) + \Big(x_s - \frac{1}{2}\Big)(\varphi_{s-1}(\mathbf{x}_s(0)) - \varphi_{s-1}(\mathbf{x}_s(1)))\end{aligned} \tag{6.20}$$

and

$$\varphi(\mathbf{x}) = \varphi_s(\mathbf{x}).$$

Then $f \in D_s^\alpha(C)$ implies that $\varphi \in D_s^\alpha(c3^s)$ and from (6.20), we have

$$\varphi_1(\mathbf{x}_1(1)) = \varphi_1(\mathbf{x}_1(0)) = \frac{f(\mathbf{x}_1(1)) + f(\mathbf{x}_1(0))}{2}$$

and

$$\int_{G_s} f(\mathbf{x})d\mathbf{x} = \int_{G_s} \varphi_1(\mathbf{x})d\mathbf{x}.$$

It is easily verified by induction that $\varphi$ satisfies (6.17) and (6.18). Hence $\varphi$ is a simple periodic function of $f$.

3. Suppose that $\psi(x) \in D_1^{\alpha+1}(c(\alpha))$ and $\psi(x)$ is a non-decreasing function in [0,1] such that

$$\psi(0) = 0, \quad \psi(1) = 1, \quad \psi'(0) = \psi'(1) = 0. \tag{6.21}$$

Then it follows from (6.21) that if $f \in D_s^\alpha(C)$, then

$$\varphi(\mathbf{x}) = f(\psi(x_1), \cdots, \psi(x_s))\psi'(x_1)\cdots\psi_s'(x_s)$$

satisfies (6.17) and (6.18) and $\varphi(\mathbf{x}) \in D_s^\alpha(C_c(\alpha)^s)$. Hence $\varphi$ is a simple periodic function of $f$. For example. take $\psi(x) = \Big(\sin\frac{\pi x}{2}\Big)^2$. Then $\psi'(x) = \frac{\pi}{2}\sin\pi x$. Obviously $\psi(\mathbf{x})$ satisfies (6.21) and

$$\varphi(\mathbf{x}) = \Big(\frac{\pi}{2}\Big)^s f\Big(\Big(\sin\frac{\pi x_1}{2}\Big)^2, \cdots, \Big(\sin\frac{\pi x_s}{2}\Big)^2\Big)\sin\pi x_1, \cdots \sin\pi x_s$$

is a simple periodic function of $f$ and $\varphi \in D_s^\alpha(C(2\pi)^{(\alpha+1)_s})$.

## 6.5 Continuation

Introduce the notation

$$\frac{\partial^l \varphi(\mathbf{x}_v(x))}{\partial x_v^l} = \frac{\partial^1 \varphi(\mathbf{x})}{\partial x_v^l}\bigg|_{x_v = x}.$$

For any $f \in D_s^\alpha(C)$, if there exists $\varphi(\mathbf{x}) \in D_s^\alpha(Cc(\alpha)^s)$ such that

$$\frac{\partial^l \varphi(\mathbf{x}_v(1))}{\partial x_v^l} = \frac{\partial^l \varphi(\mathbf{x}_v(0))}{\partial x_v^l}, \quad l = 0, 1, \cdots, \rho; v = 1, \cdots, s \tag{6.22}$$

and

$$\int_{G_s} f(\mathbf{x}) d\mathbf{x} = \int_{G_s} \varphi(\mathbf{x}) d\mathbf{x}, \tag{6.23}$$

then $\varphi(\mathbf{x})$ is called a complete periodic function of $f(\mathbf{x})$.

From (6.22),

$$\varphi(\mathbf{x}_v(1))^{(l_1,\cdots,l_s)} = \varphi(\mathbf{x}_v(0))^{(l_1,\cdots,l_s)}, \quad 0 \leqslant l_1, \cdots, l_s \leqslant \rho$$

and so by the definitions of $H_s^\alpha(C)$ and $D_s^\alpha(C)$, we derive:

**Theorem 6.5** *If $\varphi \in D_s^\alpha(C)$ and $\varphi$ is a complete periodic function of $f$, then $\varphi(\{x_1\}, \cdots, \{x_s\}) \in H_s^\alpha(C)$.*

Hence the quadrature formula for the class of functions $D_s^\alpha(C)$ can be deduced from the quadrature formula for the class of functions $H_s^\alpha(C)$. Now we shall introduce two methods of constructing the complete periodic functions.

1. Suppose that $\psi(x) \in D_1^{\alpha+1}(C)$ and $\psi(x)$ is a non-decreasing function in [0,1] such that

$$\psi(0) = 0, \psi(1) = 1, \psi^{(l)}(0) = \psi^{(l)}(1) = 0, l = 1, \cdots, \rho + 1. \tag{6.24}$$

Then if $f \in D_s^\alpha(C)$, the function

$$\varphi(\mathbf{x}) = f(\psi(x_1), \cdots, \psi(x_s))\psi'(x_1) \cdots \psi'(x_s)$$

satisfies (6.22) and (6.23) and $\varphi \in D_s^\alpha(Cc(\alpha)^s)$. Hence $\varphi$ is a complete periodic function of $f$. For example, let $n$ be an integer$\geqslant 2$ and

$$\psi_n(x) = (2n-1)C_{n-1}^{2(n-1)} \int_0^x (t(1-t))^{n-1} dt.$$

Then $\psi_n(0) = 0$ and

$$\begin{aligned}\int_0^1 t^{n-1}(1-t)^{n-1} dt &= \frac{n-1}{n} \int_0^1 t^n (1-t)^{n-2} dt = \cdots \\ &= \frac{(n-1)(n-2)\cdots 1}{n(n+1)\cdots(2n-2)} \int_0^1 t^{2n-2} dt \\ &= \frac{1}{(2n-1)C_{n-1}^{2(n-1)}}\end{aligned}$$

by integration by parts. Hence

$$\psi_n(1) = (2n-1)C_{n-1}^{2(n-1)} \int_0^1 t^{n-1}(1-t)^{n-1}dt = 1.$$

Since

$$\psi_n'(x) = (2n-1)C_{n-1}^{2(n-1)} x^{n-1}(1-x)^{n-1},$$

therefore

$$\psi_n^{(l)}(0) = \psi_n^{(l)}(1) = 0, \quad l = 1, \cdots, n-1.$$

Hence $\psi_{\rho+2}(x)$ satisfies (6.24) and so if $f \in D_s^\alpha(C)$ , then the function

$$\varphi(\mathbf{x}) = f(\psi_{\rho+2}(x_1), \cdots, \psi_{\rho+2}(x_s))\psi_{\rho+2}'(x_1)\cdots\psi_{\rho+2}'(x_s)$$

is a complete periodic function of $f$ and $\varphi \in D_s^\alpha(Cc(\alpha)^s)$. In particular, we have

$$\psi_2(x) = 6\int_0^x t(1-t)dt = 3x^2 - 2x^3$$

and

$$\psi_3(x) = 30\int_0^x t^2(1-t)^2 dt = 10x^3 - 15x^4 + 6x^5.$$

2. The rational numbers $B_n (n = 0, 1, \cdots)$ and the polynomials $B_n(x) (n = 0, 1, \cdots)$ defined by the recurrence relations

$$B_0 = 1, \quad \sum_{k=0}^{n-1} C_k^n B_k = 0, \quad n \geqslant 2 \tag{6.25}$$

and

$$B_0(x) = 1, \quad B_n(x) = \sum_{k=0}^{n} C_k^n B_k x^{n-k}, \quad n \geqslant 1 \tag{6.26}$$

are called the Bernoulli numbers and the Bernoulli polynomials respectively. For example,

$$C_0^2 B_0 + C_1^2 B_1 = 0, \quad B_1 = -\frac{1}{2},$$
$$C_0^3 B_0 + C_1^3 B_1 + C_2^3 B_2 = 0, \quad B_2 = \frac{1}{6}$$

and

$$B_1(x) = C_0^1 B_0 x + C_1^1 B_1 = x - \frac{1}{2},$$
$$B_2(x) = C_0^2 B_0 x^2 + C_1^2 B_1 x + C_2^2 B_2 = x^2 - x + \frac{1}{6}.$$

**Lemma 6.6** *Bernoulli polynomials satisfy*

$$B_n(1) - B_n(0) = \begin{cases} 1, & \text{if } n = 1, \\ 0, & \text{if } n \neq 1 \end{cases}, \tag{6.27}$$

$$B_n'(x) = nB_{n-1}(x), \quad n \geqslant 1 \tag{6.28}$$

*and*

$$\int_0^1 B_n(x)dx = 0, \quad n \geqslant 1.$$

*Proof.* Since

$$B_0(x) = 1, \quad B_1(x) = x - \frac{1}{2},$$

therefore (6.27) holds for $n = 0, 1$. Suppose that $n \geqslant 2$. Then

$$B_n(1) = \sum_{k=0}^{n} C_k^n B_k = B_n + \sum_{k=0}^{n-1} C_k^n B_k = B_n = B_n(0)$$

by (6.25) and (6.26). (6.28) may be derived immediately by the differentiation of (6.26)

$$B_n'(x) = \sum_{k=0}^{n-1} (n-k) C_k^n B_k x^{n-k-1}$$
$$= n \sum_{k=0}^{n-1} C_k^{n-1} B_k x^{n-k-1} = nB_{n-1}(x), \quad n \geqslant 1.$$

From (6.25) and (6.26), we have

$$\int_0^1 B_n(x)dx = \sum_{k=0}^{n} C_k^n B_k \frac{1}{n+1-k}$$
$$= \frac{1}{n+1} \sum_{k=0}^{n} C_k^{n+1} B_k = 0, \quad n \geqslant 1.$$

The lemma is proved.

**Lemma 6.7** *Let*

$$P_n(x) = \frac{1}{(n+1)!} B_{n+1}(x).$$

*Then*

$$P_n^{(l)}(1) - P_n^{(l)}(0) = \begin{cases} 1, & \text{if } l = n, \\ 0, & \text{if } l \neq n. \end{cases}$$

*Proof*. The lemma follow for $l > n+1$, since $B_n(x)$ is a polynomial of degree $n$. Now suppose that $l \leqslant n+1$. Then it follows from Lemma 6.6 that

$$P_n'(x) = \frac{B_{n+1}'(x)}{(n+1)!} = \frac{B_n(x)}{n!},$$

$$\cdots\cdots$$

$$P_n^{(l)}(x) = \frac{B_{n+1-l}(x)}{(n+1-l)!}$$

and

$$P_n^{(l)}(1) - P_n^{(l)}(0) = \frac{B_{n+1-l}(1) - B_{n+1-l}(0)}{(n+1-l)!} = \begin{cases} 1, & \text{if } l = n, \\ 0, & \text{if } l \neq n. \end{cases}$$

The lemma is proved.

**Lemma 6.8** *Suppose that $F(x) \in D_1^\alpha(C)$ and the function $\Phi(x)$ is defined by*

$$\Phi(x) = F(x) + \sum_{n=0}^{\rho} \sum_{\eta=0}^{1} (-1)^\eta P_n(x) F^{(n)}(\eta). \tag{6.29}$$

*Then*

$$\Phi^{(l)}(1) = \Phi^{(l)}(0), \quad l = 0, 1, \cdots, \rho.$$

*Proof*. Differentiating (6.29), we have

$$\Phi^{(l)}(x) = F^{(l)}(x) + \sum_{n=0}^{\rho} \sum_{\eta=0}^{1} (-1)^\eta P_n^{(l)}(x) F^{(n)}(\eta).$$

Hence by Lemma 6.7

$$\begin{aligned} \Phi^{(l)}(1) - \Phi^{(l)}(0) &= F^{(l)}(1) - F^{(l)}(0) \\ &\quad + \sum_{n=0}^{\rho} \sum_{\eta=0}^{1} (-1)^\eta (P_n^{(l)}(1) - P_n^{(l)}(0)) F^{(n)}(\eta) \\ &= F^{(l)}(1) - F^{(l)}(0) + \sum_{\eta=0}^{1} (-1)^\eta F^{(l)}(\eta) = 0. \end{aligned}$$

The lemma follows.

Let

$$\varphi_0(\mathbf{x}) = f(\mathbf{x})$$

and

$$\varphi_v(\mathbf{x}) = \varphi_{v-1}(\mathbf{x}) + \sum_{n_v=0}^{\rho} \sum_{\eta_v=0}^{1} (-1)^{\eta_v} P_{n_v}(x_v) \frac{\partial^{n_v} \varphi_{v-1}(\mathbf{x}_y(\eta_v))}{\partial x_v^{n_v}}. \tag{6.30}$$

for $v = 1, \cdots, s$. Let

$$\varphi(\mathbf{x}) = \varphi_s(\mathbf{x}).$$

Now we shall prove that if $f \in D_s^\alpha(C)$, then $\varphi$ is a complete periodic function of $f$.

By Lemma 6.6,

$$\int_0^1 P_n(x)dx = \frac{1}{(n+1)!}\int_0^1 B_{n+1}(x)dx = 0, \quad n \geqslant 0$$

and so by (6.30),

$$\int_{G_s} \varphi_v(\mathbf{x})d\mathbf{x} = \int_{G_s} \varphi_{v-1}(\mathbf{x})d\mathbf{x}, \quad v = 1, \cdots, s.$$

Hence

$$\int_{G_s} f(\mathbf{x})d\mathbf{x} = \int_{G_s} \varphi_1(\mathbf{x})d\mathbf{x} = \cdots$$
$$= \int_{G_s} \varphi_{s-1}(\mathbf{x})d\mathbf{x} = \int_{G_s} \varphi(\mathbf{x})d\mathbf{x}$$

and (6.23) is proved.

Let

$$F(x_v) = \varphi_{v-1}(\mathbf{x}).$$

Then by (6.29) and (6.30),

$$\Phi(x_v) = \varphi_v(\mathbf{x}).$$

Hence it follows from Lemma 6.8 that

$$\frac{\partial^{l_v}\varphi_v(\mathbf{x}_v(1))}{\partial x_v^{l_v}} = \frac{\partial^{l_v}\varphi_v(\mathbf{x}_v(0))}{\partial x_v^{l_v}}, \quad v = 1, \cdots, s;\ l_v = 0, \cdots, \rho. \tag{6.31}$$

Now we proceed to prove that

$$\frac{\partial^{l_j}\varphi_v(\mathbf{x}_j(1))}{\partial x_j^{l_j}} = \frac{\partial^{l_j}\varphi_v(\mathbf{x}_j(0))}{\partial x_j^{l_j}}, \quad j = 1, \cdots, v;\ l_j = 0, \cdots, \rho. \tag{6.32}$$

holds for $1 \leqslant v \leqslant s$. For $v = 1$, (6.32) follows from (6.31). Suppose that $v > 1$ and (6.32) holds for $v - 1$, i.e.,

$$\frac{\partial^{l_j}\varphi_{v-1}(\mathbf{x}_j(1))}{\partial x_j^{l_j}} = \frac{\partial^{l_j}\varphi_{v-1}(\mathbf{x}_j(0))}{\partial x_j^{l_j}}, \quad j = 1, \cdots, v-1;\ l_j = 0, \cdots, \rho.$$

It follows by differentiating $n_v (0 \leqslant n_v \leqslant \rho)$ times with respect to $x_v$ that

$$\frac{\partial^{l_j+n_v}\varphi_{v-1}(\mathbf{x}_j(1))}{\partial x_j^{l_j}\partial x_v^{n_v}} = \frac{\partial^{l_j+n_v}\varphi_{v-1}(\mathbf{x}_j(0))}{\partial x_j^{l_j}\partial x_v^{n_v}},$$
$$j = 1, \cdots, v-1;\ l_j = 0, \cdots, \rho.$$

and so by the differentiation of (6.30), we have

$$\frac{\partial^{l_j}\varphi_v(\mathbf{x}_j(1))}{\partial x_j^{l_j}} = \frac{\partial^{l_j}\varphi_v(\mathbf{x}_j(0))}{\partial x_j^{l_j}}, \quad j=1,\cdots,v-1;\ l_j=0,\cdots,\rho. \tag{6.33}$$

Hence (6.32) holds also for $v$ by (6.31) and (6.33). Consequently (6.32) holds for $1 \leqslant v \leqslant s$ by the induction. Especially, the case $v = s$ of (6.32) means that (6.22) holds.

$f \in D_s^\alpha(C)$ implies that $\varphi \in D_s^\alpha(Cc(\alpha)^s)$, since $B_n(x) \in H_1^\alpha(c(n))(\alpha = 1, 2, \cdots)$.

Hence we have proved that $\varphi$ is a complete periodic function of $f$.

**Notes**

The class of functions $E_s^\alpha(C)$ was introduced by N. M. Korobov [1,2,7] and the classes of functions $H_s^\alpha(C)$ and $Q_s^\alpha(C)$ were first introduced by N. S. Bahvalov [3,4] (with some modifications given by Hua Loo Keng and Wang Yuan [6,7]).

§2—§3:Cf. N. S. Bahvalov, [3,4].

§4—§5:Cf. N. M. Korobov [7] and I. F. Sarygin [1].

# Chapter 7

# Numerical Integration of Periodic Functions

## 7.1 The set of equi-distribution and numerical integration

**Theorem 7.1** *Suppose that $\alpha > 1$. Then*

$$\sup_{f\in E_s^\alpha(C)}\Big|\int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n}\sum_{l_1=0}^{m-1}\cdots\sum_{l_s=0}^{m-1} f\Big(\frac{l_1}{m},\cdots,\frac{l_s}{m}\Big)\Big| \leqslant C(2\zeta(\alpha)+1)^s n^{-\alpha/s}, \tag{7.1}$$

*where $n = m^s$ and*

$$\zeta(\alpha) = \sum_{k=1}^{\infty}\frac{1}{k^\alpha}.$$

*Proof.* The function $f$ of $E_s^\alpha(C)$ has an absolutely convergent Fourier expansion for $\alpha > 1$ (Cf. §6.1)

$$f(\mathbf{x}) = \sum C(\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})}, \quad |C(\mathbf{m})| \leqslant \frac{C}{||\mathbf{m}||^\alpha}. \tag{7.2}$$

Set $\mathbf{I} = (l_1,\cdots,l_s)$. Since

$$\sum_{k=0}^{m-1} e^{2\pi ink/m} = \begin{cases} m, & \text{if } m|n, \\ 0, & \text{if } m\dagger n, \end{cases} \tag{7.3}$$

(Cf. Lemma 3.6) and

$$C(\mathbf{0}) = \int_{G_s} f(\mathbf{x})d\mathbf{x}, \tag{7.4}$$

we have

$$\begin{aligned}\frac{1}{n}\sum_{l_1=0}^{m-1}\cdots\sum_{l_s=0}^{m-1} f\Big(\frac{l_1}{m},\cdots,\frac{l_s}{m}\Big) &= \frac{1}{n}\sum_{l_1=0}^{m-1}\cdots\sum_{l_s=0}^{m-1}\sum C(\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{l})/m}\\ &= \sum C(\mathbf{m})\prod_{j=1}^{s}\Big(\frac{1}{m}\sum_{l_j=0}^{m-1} e^{2\pi i l_j m_j/m}\Big)\\ &= \sum_{\substack{m|m_j\\ 1\leqslant j\leqslant s}} C(\mathbf{m}) = C(\mathbf{0}) + \sum_{\substack{m|m_j\\ 1\leqslant j\leqslant s}}{}' C(\mathbf{m})\end{aligned}$$

and

$$\begin{aligned}&\Big|\int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n}\sum_{l_1=0}^{m-1}\cdots\sum_{l_s=0}^{m-1} f\Big(\frac{l_1}{m},\cdots,\frac{l_s}{m}\Big)\Big|\\ &\leqslant \sum_{\substack{m|m_j\\ 1\leqslant j\leqslant s}}{}' |C(\mathbf{m})| = \sum{}' |C(m\mathbf{m})| \leqslant C\sum{}' \frac{1}{\|m\mathbf{m}\|^{\alpha}}\\ &\leqslant C\Big(\sum\frac{1}{\bar{k}^{\alpha}}\Big)^s m^{-\alpha} = C(2\zeta(\alpha)+1)^s n^{-\alpha/s}.\end{aligned}$$

The theorem is proved.

Take

$$f(\mathbf{x}) = C\frac{e^{2\pi i m x_1} + e^{-2\pi i m x_1}}{m^{\alpha}}.$$

Then $f \in E_s^{\alpha}(C)$ (of course $f \in H_s^{\alpha}(C(2\pi)^{\alpha+1})$) and

$$\begin{aligned}&\Big|\int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n}\sum_{l_1=0}^{m-1}\cdots\sum_{l_s=0}^{m-1} f\Big(\frac{l_1}{m},\cdots,\frac{l_s}{m}\Big)\Big|\\ &= \frac{2C}{m^{\alpha}} = 2Cn^{-\alpha/s},\end{aligned}$$

i.e., there exists a function of $E_s^{\alpha}(C)(\alpha > 1)$ such that the error term in the quadrature formula (7.1) is not less than $2Cn^{-\alpha/s}$. Hence the term $n^{-\alpha/n}$ in (7.1) does not admit further essential improvement.

## 7.2 The $p$ set and numerical integration

**Theorem 7.2** Suppose that $\alpha > 1$ and $n = p^2$. Then

$$\begin{aligned}\sup_{f\in E_s^{\alpha}(C)}\Big|\int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n}\sum_{a=1}^{p}\sum_{k=1}^{p} f\Big(\frac{a}{p},\frac{ak}{p},\cdots,\frac{ak^{s-1}}{p}\Big)\Big| \\ \leqslant Cs(2\zeta(\alpha)+1)^s n^{-\frac{1}{2}}.\end{aligned} \tag{7.5}$$

*Proof*. Set $\mathbf{k} = (1, k, \cdots, k^{s-1})$. Then by (7.2),(7.3) and (7.4),

$$\sum_{a=1}^{p}\sum_{k=1}^{p} f\Big(\frac{a}{p}, \frac{ak}{p}, \cdots, \frac{ak^{s-1}}{p}\Big) = \sum C(\mathbf{m}) \sum_{a=1}^{p}\sum_{k=1}^{p} e^{2\pi i(\mathbf{k},\mathbf{m})a/p}$$
$$= p^2 C(\mathbf{0}) + p{\sum}' C(\mathbf{m}) \sum_{\substack{1\leqslant k\leqslant p \\ (\mathbf{k},\mathbf{m})\equiv 0(\bmod p)}} 1$$

and so

$$\sup_{f\in E_s^{\alpha}(C)} \Big| \int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n}\sum_{a=1}^{p}\sum_{k=1}^{p} f\Big(\frac{a}{p}, \frac{ak}{p}, \cdots, \frac{ak^{s-1}}{p}\Big)\Big|$$
$$\leqslant \frac{C}{p}{\sum}' \frac{1}{||\mathbf{m}||^{\alpha}} \sum_{\substack{1\leqslant k\leqslant p \\ (\mathbf{k},\mathbf{m})\equiv 0(\bmod p)}} 1.$$

Since the number of solutions of the congruence

$$(\mathbf{k}, \mathbf{m}) \equiv 0 (\bmod\ p), \quad 1 \leqslant k \leqslant p$$

is at most s−1, if $m_1, \cdots, m_s$ are not all divisible by $p$, and is equal to $p$ otherwise (Cf. Lemma 4.5), we have

$${\sum}' \frac{1}{||\mathbf{m}||} \sum_{\substack{1\leqslant k\leqslant p \\ (\mathbf{k},\mathbf{m})\equiv 0(\bmod p)}} 1 \leqslant (s-1){\sum}' \frac{1}{||\mathbf{m}||^{\alpha}} + p{\sum}' \frac{1}{||p\mathbf{m}||^{\alpha}}$$
$$\leqslant s\sum \frac{1}{||\mathbf{m}||^{\alpha}} = s\Big(\sum \frac{1}{m^{\alpha}}\Big)^s = s(2\zeta(\alpha)+1)^s.$$

The theorem follows.

**Lemma 7.1** *The number of non-negative integral solutions of*

$$n = r_1 + \cdots, + r_s$$

*is equal to* $C_{s-1}^{n+s-1}$.

*Proof*. For $|x| < 1$,

$$\frac{1}{(1-x)^s} = \Big(\sum_{r=0}^{\infty} x^r\Big)^s = \sum_{\substack{r_1=0 \\ n=r_1+\cdots+r_s}}^{\infty} \cdots \sum_{r_s=0}^{\infty} x^n$$

and so

$$\frac{1}{(1-x)^s} = \sum_{n=0}^{\infty} C_{s-1}^{n+s-1} x^n.$$

The lemma follows by comparing the coefficients of $x^n$ in the above two formulas.

**Lemma 7.2** *Suppose that $s$ is an integer* $\geqslant 0$ and $\alpha > 0$. *Then*

$$\sum_{t=n}^{\infty} t^s 2^{-\alpha t} \leqslant c(\alpha, s) n^s 2^{-\alpha n}.$$

*Proof*. Since

$$t^s 2^{-\alpha t} \leqslant \int_t^{t+1} u^s 2^{-\alpha(u-1)} du, \quad t = n, n+1, \cdots,$$

therefore

$$\begin{aligned}\sum_{t=n}^{\infty} t^s 2^{-\alpha t} &\leqslant \int_n^{\infty} u^s 2^{-\alpha(u-1)} du \\ &= \frac{n^s 2^{-\alpha(n-1)}}{\alpha \ln 2} + \frac{s}{\alpha \ln 2} \int_n^{\infty} u^{s-1} 2^{-\alpha(u-1)} du \\ &= \cdots \\ &= \sum_{k=0}^{s} \frac{k! C_k^s n^{s-k} 2^{-\alpha(n-1)}}{(\alpha \ln 2)^{k+1}} \leqslant c(\alpha, s) n^s 2^{-\alpha n}.\end{aligned}$$

The lemma is proved.

**Theorem 7.3** *Suppose that* $0 \leqslant \alpha < 1$. *Then*

$$\sup_{f \in Q_s^\alpha(G)} \Big| \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{p} \sum_{k=1}^{p} f\Big(\frac{k}{p}, \frac{k^2}{p}, \cdots, \frac{k^s}{p}\Big) \Big|$$
$$\leqslant \begin{cases} Cc(\alpha, s) p^{-\frac{1}{2}}, \text{ if } \dfrac{1}{2} < \alpha \leqslant 1, \\ Cc(\alpha, s) p^{-\alpha} (\ln p)^{s-1+\delta_{\frac{1}{2},\alpha}}, \text{ if } 0 < \alpha \leqslant \dfrac{1}{2}. \end{cases}$$

*Proof*. For $f \in Q_s^\alpha(C)$, let

$$S(f) = \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{p} \sum_{k=1}^{p} f\Big(\frac{k}{p}, \frac{k^2}{p}, \cdots, \frac{k^s}{p}\Big).$$

Then by Theorem 6.3, we have

$$S(f) = \sum\nolimits'' S(\varphi_t),$$

where

$$S(\varphi_t) = \int_{G_s} \varphi_t(\mathbf{x}) d\mathbf{x} - \frac{1}{p} \sum_{k=1}^{p} \varphi_t\Big(\frac{k}{p}, \frac{k^2}{p}, \cdots, \frac{k^s}{p}\Big).$$

Hence

$$\sup_{f \in Q_s^\alpha(C)} |S(f)| \leqslant \sum\nolimits_1 + \sum\nolimits_2, \tag{7.6}$$

where

$$\sum\nolimits_1 = \sup_{f\in Q_s^\alpha(C)} \sum_{t_0\geqslant \log_2 p}'' |S(\varphi_{\mathbf{t}})|$$

and

$$\sum\nolimits_2 = \sup_{f\in Q_s^\alpha(C)} \sum_{t_0< \log_2 p}'' |S(\varphi_{\mathbf{t}})|,$$

in which $t_0 = t_1 + \cdots + t_s$. Since $C_{s-1}^{t+s-1} \leqslant c(s)t^{s-1}$ and

$$|S(\varphi_{\mathbf{t}})| \leqslant 2||\varphi_{\mathbf{t}}|| \leqslant C2^{1-\alpha t_0},$$

therefore

$$\begin{aligned}\sum\nolimits_1 &\leqslant Cc(s) \sum_{t_0\geqslant \log_2 p}'' 2^{-\alpha t_0} \leqslant Cc(s) \sum_{t=[\log_2 p]}^{\infty} t^{s-1}2^{-\alpha t} \\ &\leqslant Cc(\alpha, s)p^{-\alpha}(\ln p)^{s-1}\end{aligned} \tag{7.7}$$

by Lemmas 7.1. and 7.2. Set $\mathbf{k} = (k, k^2, \cdots, k^s)$. Since

$$C_{\mathbf{t}}(0) = \int_{G_s} \varphi_{\mathbf{t}}(\mathbf{x})d\mathbf{x},$$

we have

$$\begin{aligned}\sum_{k=1}^{p} \varphi_{\mathbf{t}}\left(\frac{k}{p}, \frac{k^2}{p}, \cdots, \frac{k^s}{p}\right) &= \sum_{k=1}^{p} \sum C_{\mathbf{t}}(\mathbf{m})e^{2\pi i(\mathbf{k},\mathbf{m})/p} \\ &= pC_{\mathbf{t}}(\mathbf{0}) + \sum\nolimits' C_{\mathbf{t}}(\mathbf{m}) \sum_{k=1}^{p} e^{2\pi i(\mathbf{k},\mathbf{m})/p}\end{aligned}$$

and

$$|S(\varphi_{\mathbf{t}})| \leqslant \frac{1}{p}\sum\nolimits' |C_{\mathbf{t}}(\mathbf{m})| \Big|\sum_{k=1}^{p} e^{2\pi i(\mathbf{k},\mathbf{m})/p}\Big|.$$

Suppose that $\mathbf{m} \neq \mathbf{0}$. If at least one of the relations $|m_i| < 2^{t_i} (1 \leqslant i \leqslant s)$ is not satisfied, then $C_{\mathbf{t}}(\mathbf{m})$=0. Otherwise if $|m_i| < 2^{t_i} (1 \leqslant i \leqslant s)$ and $t_0 < \log_2 p$, then $m_i (1 \leqslant i \leqslant s)$ are not all divisible by $p$. Hence

$$\sum\nolimits_2 \leqslant (s-1)p^{-\frac{1}{2}} \sup_{f\in Q_s^\alpha(C)} \sum_{t_0<\log_2 p}'' \sum_{\substack{|m_i|<2^{t_i} \\ 1\leqslant i\leqslant s}}' |C_{\mathbf{t}}(\mathbf{m})|$$

by Lemma 4.7. Let

$$||F||_2 = \left(\int_{G_s} |F(\mathbf{x})|^2 d\mathbf{x}\right)^{1/2}.$$

Then

$$\sum |G_{\mathbf{t}}(\mathbf{m})|^2 = ||\varphi_{\mathbf{t}}||_2^2 \leqslant ||\varphi_{\mathbf{t}}||^2 \leqslant C^2c(\alpha, s)2^{-2\alpha t_0}$$

and so by Lemmas 7.1 and 7.2 and Schwarz's inequality,

$$\begin{aligned}
\sum\nolimits_2 &\leqslant (s-1)p^{-\frac{1}{2}} \sup_{f\in Q_s^\alpha(C)} \sum_{t_0<\log_2 p}'' \Big(\sum_{|m_i|<2^{t_i}}' 1\Big)^{1/2} \Big(\sum_{|m_i|<2^{t_i}}' |C_{\mathbf{t}}(\mathbf{m})|^2\Big)^{1/2} \\
&\leqslant (s-1)p^{-\frac{1}{2}} \sup_{f\in Q_s^\alpha(C)} \sum_{t_0<\log_2 p}'' 2^{\frac{s+t_0}{2}} \|\varphi_{\mathbf{t}}\|_2 \\
&\leqslant Cc(\alpha,s)p^{-\frac{1}{2}} \sum_{t_0<\log_2 p}'' 2^{-(\alpha-\frac{1}{2})t_0} \\
&\leqslant Cc(\alpha,s)p^{-\frac{1}{2}} \sum_{t=0}^{[\log_2 p]} (t+1)^{s-1} 2^{-(\alpha-\frac{1}{2})t} \\
&\leqslant \begin{cases} Cc(\alpha,s)p^{-\frac{1}{2}}, \text{ if } \frac{1}{2}<\alpha\leqslant 1, \\ Cc(\alpha,s)p^{-\alpha}(\ln p)^{s-1+\delta_{\frac{1}{2},\alpha}}, \text{ if } 0<\alpha\leqslant \frac{1}{2}. \end{cases}
\end{aligned} \tag{7.8}$$

The theorem follows from (7.6), (7.7) and (7.8).

For $\alpha > 1$, we may also use the sets of points

$$\left(\left\{\frac{k}{p^2}\right\}, \left\{\frac{k^2}{p^2}\right\}, \cdots, \left\{\frac{k^s}{p^2}\right\}\right), \quad 1\leqslant k\leqslant p^2$$

and

$$\left(\left\{\frac{k}{p}\right\}, \left\{\frac{k^2}{p}\right\}, \cdots, \left\{\frac{k^s}{p}\right\}\right), \quad 1\leqslant k\leqslant p$$

(Cf. §4.3) to obtain the quadrature formulas for the class of functions $E_s^\alpha(C)$ which have the same precision as Theorem 7.2.

Now, we shall study the lower estimate of the error term of the quadrature formula given by Theorem 7.2. Set

$$f(\mathbf{x}) = C \sum_{m_1=-(p-1)}^{p-1}{}' \sum_{m_2=-(p-1)}^{p-1}{}' \frac{e^{2\pi i(m_1x_1+m_2x_2)}}{(\bar{m}_1\bar{m}_2)^\alpha}, \quad \alpha>0.$$

Then $f \in E_s^\alpha(C)$,

$$\int_{G_s} f(\mathbf{x})d\mathbf{x} = 0$$

and

$$\frac{1}{p^2}\sum_{a=1}^{p}\sum_{k=1}^{p} f\Big(\frac{a}{p}, \frac{ak}{p}, \cdots, \frac{ak^{s-1}}{p}\Big)$$
$$= \frac{C}{p^2}\sum_{a=1}^{p}\sum_{k=1}^{p}\ \sideset{}{'}\sum_{m_1=-(p-1)}^{p-1}\ \sideset{}{'}\sum_{m_2=-(p-1)}^{p-1} \frac{e^{2\pi i(m_1+m_2k)_{a/p}}}{(\bar{m}_1\bar{m}_2)^\alpha}$$
$$= \frac{C}{p}\sum_{k=1}^{p}\ \sideset{}{'}\sum_{m_1=-(p-1)}^{p-1}\ \sideset{}{'}\sum_{\substack{m_2=-(p-1)\\ m_1+m_2k\equiv 0 (\mathrm{mod}\ p)}}^{p-1} \frac{1}{(\bar{m}_1\bar{m}_2)^\alpha} \geqslant Cp^{-1}$$

Hence there exists a function of $E_s^\alpha(C)$ such that the error term in quadrature formula (7.5) is not less than $Cn^{-1/2}$ and so the error term in (7.5) does not admit further essential improvement.

## 7.3 The *gp* set and numerical integration

**Lemma 7.3** *Suppose that $\alpha \geqslant 1$ and $a_i \geqslant 0 (-\infty < i < \infty)$. If*

$$\sum a_i < \infty,$$

*then*

$$\sum a_i^\alpha \leqslant \big(\sum a_i\big)^\alpha.$$

*Proof.* Clearly, the lemma is true if all $a_i = 0$. Now suppose that $\sum a_i > 0$. Then

$$\sum a_i^\alpha = \sum_i \Big(\frac{a_i}{\sum_j a_j}\Big)^\alpha \Big(\sum_k a_k\Big)^\alpha.$$
$$\leqslant \Big(\sum_k a_k\Big)^\alpha \sum_i \frac{a_i}{\sum\limits_j a_j} \leqslant \big(\sum a_k\big)^\alpha$$

The lemma is proved.

Let $\alpha$ be a positive number and $l$ be the least integer$\geqslant \alpha$. Let $\mu_{n,l,k}$ be the set of integers defined by

$$\Big(\sum_{k=-n}^{n} z^k\Big)^l = \sum_{k=-nl}^{nl} \mu_{n,lk} z^k.$$

**Theorem 7.4** *Suppose that $\alpha > 1$. If $\gamma$ is a real vector such that*

$$\langle(\mathbf{m}, \gamma)\rangle > b\|\mathbf{m}\|^{-a}$$

*holds for all integral vectors* $\mathbf{m} \neq \mathbf{0}$, *where* $a$, $b$ *are two constants satisfying* $s \geqslant a \geqslant 1$ *and* $b > 0$, *then*

$$\sup_{f \in E_s^\alpha(C)} \Big| \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{(2n+1)^l} \sum_{k=-nl}^{nl} \mu_{n,l,k} f(k\boldsymbol{\gamma}) \Big| \leqslant Cc(b, \alpha, s) n^{-\alpha + \frac{s\alpha^2(a-1)}{\alpha-1}} (\ln n)^{\alpha + s\alpha\delta_{1,a}}.$$

*Proof.* For $f \in E_s^\alpha(C)$, we have

$$\begin{aligned}
&\frac{1}{(2n+1)^l} \sum_{k=-nl}^{nl} \mu_{n,l,k} f(k\boldsymbol{\gamma}) \\
&= \frac{1}{(2n+1)^l} \sum C(\mathbf{m}) \sum_{k=-nl}^{nl} \mu_{n,l,k} e^{2\pi i (\mathbf{m}, \boldsymbol{\gamma}) k} \\
&= C(\mathbf{0}) + \frac{1}{(2n+1)^l} \sum{}' C(\mathbf{m}) \Big( \sum_{k=-n}^{n} e^{2\pi i (\mathbf{m}, \boldsymbol{\gamma})^k} \Big)^l
\end{aligned}$$

and so

$$\begin{aligned}
&\sup_{f \in E_s^\alpha(C)} \Big| \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{(2n+1)^l} \sum_{k=-nl}^{nl} \mu_{n,l,k} f(k\boldsymbol{\gamma}) \Big| \\
&\leqslant C \sum{}' \frac{1}{||\mathbf{m}||^\alpha} \Big| \frac{1}{2n+1} \sum_{k=-n}^{n} e^{2\pi i (\mathbf{m}, \boldsymbol{\gamma})^k} \Big|^l \qquad (7.9)\\
&= C(\sum\nolimits_1 + \sum\nolimits_2),
\end{aligned}$$

where $\sum_1$ denotes a sum of $\mathbf{m}$ satisfying $|m_i| \leqslant n^{\frac{\alpha}{\alpha-1}} (1 \leqslant i \leqslant s)$ and $\sum_2$ the remaining part. Since

$$\Big| \sum_{k=-n}^{n} e^{2\pi i (\mathbf{m}, \boldsymbol{\gamma})^k} \Big| \leqslant \min\Big(2n+1, \frac{1}{2\langle(\mathbf{m}, \boldsymbol{\gamma})\rangle}\Big)$$

by Lemma 3.10, we have

$$\begin{aligned}
\sum\nolimits_1 &\leqslant \frac{1}{2^\alpha (2n+1)^\alpha} \sum_{|m_i| \leqslant n^{\frac{\alpha}{\alpha-1}}}{}' \frac{1}{||\mathbf{m}||^\alpha \langle(\mathbf{m}, \boldsymbol{\gamma})\rangle^\alpha} \\
&\leqslant c(\alpha) n^{-\alpha} \Big( \sum_{|m_i| \leqslant n^{\frac{\alpha}{\alpha-1}}}{}' \frac{1}{||\mathbf{m}|| \langle(\mathbf{m}, \boldsymbol{\gamma})\rangle} \Big)^\alpha \qquad (7.10)\\
&\leqslant c(b, \alpha, s) n^{-\alpha + \frac{s\alpha^2(a-1)}{a-1}} (\ln n)^{\alpha + s\alpha\delta_{1,a}}
\end{aligned}$$

by Lemmas 3.14 and 7.3. Since

$$\sum_{k>n} \frac{1}{k^\alpha} \int_n^\infty \frac{dt}{t^\alpha} = \frac{1}{\alpha-1} n^{-(\alpha-1)},$$

therefore

$$\sum\nolimits_2 \leqslant \sum_{i=1}^{s} \sum_{|m_i|>n^{\frac{\alpha}{\alpha-1}}} \frac{1}{|m_i|^\alpha} \sum \frac{1}{||\mathbf{m}||^\alpha} \leqslant c(\alpha, s) n^{-\alpha}. \tag{7.11}$$

The theorem follows from (7.9), (7.10) and (7.11).

**Theorem 7.5** *Suppose that* $0 < \alpha \leqslant 1$. *Then under the assumption of Theorem* 7.4,

$$\sup_{f \in Q_s^\alpha(C)} \left| \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{k=1}^{n} f(k\gamma) \right| \leqslant Cc(b, \alpha, s) n^{-\alpha+s(a-1)} (\ln\, n)^{s-1+s\delta_{1,a}}.$$

*Proof.* For given $f \in Q_s^\alpha(C)$, let

$$S(f) = \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{k=1}^{n} f(k\gamma).$$

Then by Theorem 6.3,

$$S(f) = \sum\nolimits'' S(\varphi_{\mathbf{t}}),$$

where

$$S(\varphi_{\mathbf{t}}) = \int_{G_s} \varphi_{\mathbf{t}}(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{k=1}^{n} \varphi_{\mathbf{t}}(k\gamma).$$

Hence

$$\sup_{f \in Q_s^\alpha(C)} |S(f)| \leqslant \sum\nolimits_1 + \sum\nolimits_2, \tag{7.12}$$

where

$$\sum\nolimits_1 = \sup_{f \in Q_s^\alpha(C)} \sum_{t_0 \geqslant \log_2 n}'' |S(\varphi_{\mathbf{t}})|$$

and

$$\sum\nolimits_2 = \sup_{f \in Q_s^\alpha(C)} \sum_{t_0 < \log_2 n}'' |S(\varphi_{\mathbf{t}})|.$$

Similar to (7.7), we have

$$\sum\nolimits_1 \leqslant Cc(s) \sum_{t_0 \geqslant \log_2 n}'' 2^{-\alpha t_0} \leqslant Cc(\alpha, s) n^{-\alpha} (\ln\, n)^{s-1}. \tag{7.13}$$

Since $C_{\mathbf{t}}(\mathbf{m}) = 0$ for $||\mathbf{m}|| \geqslant 2^{t_0}$, therefore by Lemma 3.10,

$$\begin{aligned} \sum_{k=1}^{n} \varphi_{\mathbf{t}}(k\gamma) &= \sum_{k=1}^{n} \sum_{||\mathbf{m}||<2^{t_0}} C_{\mathbf{t}}(\mathbf{m}) e^{2\pi i (\mathbf{m}, \gamma) k} \\ &= nC_{\mathbf{t}}(\mathbf{0}) + \sum_{||\mathbf{m}||<2^{t_0}}' C_{\mathbf{t}}(\mathbf{m}) \sum_{k=1}^{n} e^{2\pi i (\mathbf{m}, \gamma) k} \end{aligned}$$

and so

$$|S(\varphi_{\mathbf{t}})| \leqslant n^{-1} \sum_{||\mathbf{m}||<2^{t_0}}' |C_{\mathbf{t}}(\mathbf{m})| \frac{\mathbf{1}}{\mathbf{2}\langle(\mathbf{m},\gamma)\rangle}.$$

Hence it follows by (6.16), Theorem 6.2 and Lemma 3.14 that

$$\begin{aligned}
\sum\nolimits_2 &\leqslant \sup_{f\in Q_s^\alpha(C)} n^{-1} \sum_{||\mathbf{m}||<n}' \frac{1}{\langle(\mathbf{m},\gamma)\rangle} \sum\nolimits'' |C_{\mathbf{t}}(\mathbf{m})| \\
&\leqslant \sup_{f\in Q_s^\alpha(C)} n^{-1} \sum_{||\mathbf{m}||<n}' \frac{|C(\mathbf{m})|}{\langle(\mathbf{m},\gamma)\rangle} \\
&\leqslant Cc(\alpha,s) n^{-1} \sum_{||\mathbf{m}||<n}' \frac{1}{||\mathbf{m}||^\alpha \langle(\mathbf{m},\gamma)\rangle} \\
&\leqslant Cc(\alpha,s) n^{-1} \sum_{||\mathbf{m}||<n}' \frac{n^{1-\alpha}}{||\mathbf{m}|| \langle(\mathbf{m},\gamma)\rangle} \\
&\leqslant Cc(b,\alpha,s) n^{-\alpha+s(a-1)} (\ln n)^{1+s\delta_{1,a}}.
\end{aligned} \tag{7.14}$$

The theorem follows by (7.12), (7.13) and (7.14).

For $\alpha = 2$, the weight $\mu_{n,l,k}$ may be simplified. First, we shall state the following lemma.

**Lemma 7.4** *Let $\delta$ be a real number. Then*

$$\Big| \sum_{k=-(n-1)}^{n-1} (n-|k|) e^{2\pi i k\delta} \Big| \leqslant \min\Big(n^2, \frac{1}{4\langle\delta\rangle^2}\Big).$$

*Proof.* Since

$$\sum_{k=-(n-1)}^{n-1} (n-|k|) = \sum_{j=0}^{n-1} \sum_{k=-j}^{j} 1 = n^2$$

and $\sin \pi\delta \geqslant 2\delta$ for $0 \leqslant \delta \leqslant \frac{1}{2}$, so if $\delta$ is not an integer, then

$$\begin{aligned}
\sum_{k=-(n-1)}^{n-1} (n-|k|) e^{2\pi i k\delta} &= \sum_{j=0}^{n-1} \sum_{k=-j}^{j} e^{2\pi i k\delta} \\
&= \frac{1}{\sin\pi\delta} \sum_{j=0}^{n-1} \sin(2j+1)\pi\delta = \Big(\frac{\sin n\pi\delta}{\sin \pi\delta}\Big)^2 \leqslant \frac{1}{4\langle\delta\rangle^2}.
\end{aligned}$$

The lemma is proved.

**Theorem 7.6** *Under the assumption of Theorem 7.4,*

$$\begin{aligned}
\sup_{f\in E_s^2(C)} \Big| \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{k=-(n-1)}^{n-1} \Big(1 - \frac{|k|}{n}\Big) f(k\gamma) \Big| \\
\leqslant Cc(b,s) n^{-2+4s(a-1)} (\ln n)^{2+2s\delta_{1,a}}.
\end{aligned}$$

*Proof.* Since

$$\begin{aligned}
&\frac{1}{n}\sum_{k=-(n-1)}^{n-1}\left(1-\frac{|k|}{n}\right)f(k\gamma)\\
&=\frac{1}{n^2}\sum_{k=-(n-1)}^{n-1}(n-|k|)\sum C(\mathbf{m})e^{2\pi i(\mathbf{m},\gamma)}k\\
&=\frac{1}{n^2}\sum C(\mathbf{m})\sum_{k=-(n-1)}^{n-1}(n-|k|)e^{2\pi i(\mathbf{m},\gamma)}k\\
&=C(\mathbf{0})+\sum{}' C(\mathbf{m})\frac{1}{n^2}\sum_{k=-(n-1)}^{n-1}(n-|k|)e^{2\pi i(\mathbf{m},\gamma)}k,
\end{aligned}$$

so by Lemma 7.4,

$$\begin{aligned}
&\sup_{f\in E_s^2(C)}\left|\int_{G_s}f(\mathbf{x})d\mathbf{x}-\frac{1}{n}\sum_{k=-(n-1)}^{n-1}\left(1-\frac{|k|}{n}\right)f(k\gamma)\right|\\
&\leqslant Cn^{-2}\sum{}'\frac{1}{\|\mathbf{m}\|^2}\min\left(\frac{1}{4\langle(\mathbf{m},\gamma)\rangle^2},n^2\right)
\end{aligned}$$

and so the theorem may be easily proved by a method similar to the proof of Theorem 7.4.

Using the notation of §4.5, it follows by Theorems 4.12, 4.13 and 7.6 that

**Theorem 7.7** *We have*

$$\begin{aligned}
&\sup_{f\in E_s^2(C)}\left|\int_{G_s}f(\mathbf{x})d\mathbf{x}-\frac{1}{n}\sum_{k=-(n-1)}^{n-1}\left(1-\frac{|k|}{n}\right)f(k\alpha)\right|\\
&\leqslant Cc(\alpha,\varepsilon)n^{-2+\varepsilon}
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{f\in E_s^2(C)}\left|\int_{G_s}f(\mathbf{x})d\mathbf{x}-\frac{1}{n}\sum_{k=-(n-1)}^{n-1}\left(1-\frac{|k|}{n}\right)f(k\beta)\right|\\
&\leqslant Cc(\beta,\varepsilon)n^{-2+\varepsilon}.
\end{aligned}$$

From Theorems 4.12, 4.13, 7.4 and 7.5 we may obtain similar quadrature formulas.

## 7.4 The lower estimation of the error term for the quadrature formula

**Theorem 7.8** *For any given sequence of* $G_s$

$$P(k)=(x_1(k),\cdots,x_s(k)),\quad k=1,2,\cdots,$$

*the error term of the quadrature formula*

$$\int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n}\sum_{k=1}^{n} f(P(k))$$

*for the class of analytic functions on $G_s$ can not be better than $O(n^{-1})$, where the constant implied by the symbolO depends on $f$ only.*

*Proof.* Suppose that the theorem is not true. Then there exists a sequence $P(k)(k = 1, 2, \cdots)$ such that

$$\int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n}\sum_{k=1}^{n} f(P(k)) = o(n^{-1})$$

holds for any analytic function $f$, where the constant implied by the symbol $o$ depends on $f$ only. For example, if we take $f(\mathbf{x}) = g(x_1)$, then

$$\sum_{k=1}^{n} g(x_1(k)) = n\int_0^1 g(x)dx + o(1).$$

Hence

$$g(x_1(n)) = \sum_{k=1}^{n} g(x_1(k)) - \sum_{k=1}^{n-1} g(x_1(k)) = \int_0^1 g(x)dx + o(1).$$

In particular, we have

$$\sin(2\pi x_1(n)) = o(1)$$

and

$$\cos(2\pi x_1(n)) = o(1),$$

if we take $g(x) = \sin 2\pi x$ and $g(x) = \cos 2\pi x$ respectively. Hence

$$1 = (\sin(2\pi x_1(n)))^2 + (\cos(2\pi x_1(n)))^2 = o(1).$$

This leads to a contradiction. The theorem follows.

From Theorem 7.8 we know that for any given sequence of points $P(k)(k = 1, 2, \cdots)$ of $G_s$, if we use the simple sum

$$n^{-1}\sum_{k=1}^{n} f(P(k))$$

to approximate the definite integral on $G_s$, then the error term is comparatively large. The quadrature formula given by Theorem 7.4 is quite precise but its disadvantage is the complicated weight $\mu_{n,l,k}$ which depends on $\alpha$.

## 7.5 The solutions of congruences and numerical integration

**Theorem 7.9** *Suppose that* $\alpha > 1$. *If the congruence*

$$(\mathbf{a}, \mathbf{m}) \equiv 0 (\bmod\ n) \tag{7.15}$$

*has no solution in the domain*

$$\|\mathbf{m}\| \leqslant M, \quad \mathbf{m} \neq \mathbf{0},$$

*then*

$$\sup_{f \in E_s^\alpha(C)} \left| \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) \right| \leqslant Cc(\alpha, \varepsilon)^s M^{-\alpha+\varepsilon}. \tag{7.16}$$

*Proof.* Obviously, we may suppose that $\varepsilon < \alpha - 1$. Since

$$\begin{aligned}
\frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) &= \frac{1}{n} \sum_{k=1}^{n} \sum C(\mathbf{m}) e^{2\pi i (\mathbf{a}, \mathbf{m})} k/n \\
&= C(\mathbf{0}) + \sum{}' C(\mathbf{m}) \frac{1}{n} \sum_{k=1}^{n} e^{2\pi i (\mathbf{a}, \mathbf{m})} k/n \\
&= C(\mathbf{0}) + \sum_{(\mathbf{a}, \mathbf{m}) \equiv 0 (\bmod\ n)}{}' C(\mathbf{m}),
\end{aligned}$$

we have

$$\sup_{f \in E_s^\alpha(C)} \left| \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) \right| \leqslant C \sum_{(\mathbf{a}, \mathbf{m}) \equiv 0 (\bmod\ n)}{}' \frac{1}{\|\mathbf{m}\|^\alpha}.$$

Let $T_{l,M}$ be the number of solutions of (7.15) in the domain

$$\|\mathbf{m}\| < lM$$

where $l$ is a positive integer. Then

$$T_{l,M} \leqslant c(\varepsilon)^l l^{1+\varepsilon} M^\varepsilon$$

by Lemma 3.3. Since

$$l^{-\alpha} - (1 + l)^{-\alpha} = \alpha \int_l^{l+1} x^{-\alpha-1} dx \leqslant \alpha l^{-(\alpha+1)},$$

therefore

$$\begin{aligned}\sum_{(\mathbf{a},\mathbf{m})\equiv 0(\bmod\ n)}{}' \frac{1}{\|\mathbf{m}\|^\alpha} &\leqslant \sum_{l=1}^{\infty} \frac{T_{l+1,M}-T_{l,M}}{(lM)^\alpha}\\ &= \sum_{l=1}^{\infty} T_{l+1,M}(l^{-\alpha}-(l+1)^{-\alpha})M^{-\alpha}\\ &\leqslant c(\alpha,\varepsilon)^s M^{-\alpha+\varepsilon}\sum_{l=1}^{\infty} l^{-\alpha+\varepsilon} \leqslant c(\alpha,\varepsilon)^s M^{-\alpha+\varepsilon}.\end{aligned}$$

The theorem follows.

**Theorem 7.10** *Suppose that* $0<\alpha\leqslant 1$. *Then under the assumption of Theorem 7.9,*

$$\sup_{f\in Q_s^\alpha(C)}\left|\int_{G_s} f(\mathbf{x})d\mathbf{x}-\frac{1}{n}\sum_{k=1}^{n} f\Big(\frac{k\mathbf{a}}{n}\Big)\right| \leqslant Cc(\alpha,\varepsilon)^s M^{-\alpha+\varepsilon}. \tag{7.17}$$

*Proof.* We may suppose that $\varepsilon<\alpha$. Let

$$S(f)=\int_{G_s} f(\mathbf{x})d\mathbf{x}-\frac{1}{n}\sum_{k=1}^{n} f\Big(\frac{k\mathbf{a}}{n}\Big).$$

Then

$$S(f)=\sum{}'' S(\varphi_{\mathbf{t}}),$$

where

$$S(\varphi_{\mathbf{t}})=\int_{G_s}\varphi_{\mathbf{t}}(\mathbf{x})d\mathbf{x}-\frac{1}{n}\sum_{k=1}^{n}\varphi_{\mathbf{t}}\Big(\frac{k\mathbf{a}}{n}\Big).$$

Hence

$$\sup_{f\in Q_s^\alpha(C)}|S(f)|\leqslant \sum\nolimits_1+\sum\nolimits_2,$$

where

$$\sum\nolimits_1=\sup_{f\in Q_s^\alpha(C)}\sum_{t_0\geqslant \log_2 M}{}'' |S(\varphi_{\mathbf{t}})|$$

and

$$\sum\nolimits_2=\sup_{f\in Q_s^\alpha(C)}\sum_{t_0< \log_2 M}{}'' |S(\varphi_{\mathbf{t}})|.$$

Similar to (7.7), we have

$$\begin{aligned}\sum\nolimits_1 &\leqslant 2C\sum_{t_0\geqslant\log_2 M}{}'' 2^{-\alpha t_0}=2C\sum_{t_0\geqslant\log_2 M}{}'' 2^{-(\alpha-\varepsilon)t_0-\varepsilon t_0}\\ &\leqslant 2CM^{-\alpha+\varepsilon}\Big(\sum_{t=0}^{\infty}2^{-\varepsilon t}\Big)^s = Cc(\alpha,\varepsilon)^s M^{-\alpha+\varepsilon}.\end{aligned}$$

Since $C_{\mathbf{t}}(\mathbf{m}) = 0$ for $||\mathbf{m}|| \geqslant 2^{t_0}$ and

$$|S(\varphi_{\mathbf{t}})| = \Big|\sum{}' C_{\mathbf{t}}(\mathbf{m}) \frac{1}{n} \sum_{k=1}^{n} e^{2\pi i (\mathbf{a},\mathbf{m})} k/n\Big| \leqslant \sum_{(\mathbf{a},\mathbf{m})\equiv 0 (\mathrm{mod}\ n)}{}' |C_{\mathbf{t}}(\mathbf{m})|,$$

hence

$$\sum{}_2 \leqslant \sup_{f\in Q_s^\alpha(C)} \sum_{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ n)}{}' \sum_{t_0<\log_2 M}{}'' |C_{\mathbf{t}}(\mathbf{m})| = 0$$

The theorem follows.

If we use Lemma 3.5 instead of Lemma 3.3 in the proof of Theorem 7.9 and the Lemmas 7.1 and 7.2 to evaluate $\sum_1$ in the proof of Theorem 7.10, then we have

**Theorem 7.11** *Under the assumptions of Theorems* 7.9 *and* 7.10, *the right hand sides of* (7.16) *and* (7.17) *may be replaced by*

$$Cc(\alpha, s)M^{-\alpha}(\ln 3M)^{s-1}.$$

Using the notations of §4.6 and §4.7, we have the following two lemmas by (1.14), Theorems 4.12 and 2.8 and Lemma 3.9.

**Lemma 7.5** *The congruence*

$$c_1m_1 + c_2m_2 + \cdots + c_sm_s \equiv 0(\mathrm{mod}\ n), \quad s = \frac{\varphi(m)}{2}$$

*has no solution in the domain*

$$||\mathbf{m}|| \leqslant c(\mathscr{R}_s, \varepsilon)n^{\frac{1}{2}+\frac{1}{2(s-1)}-\varepsilon}, \quad \mathbf{m} \neq \mathbf{0}.$$

**Lemma 7.6** *The congruence*

$$m_1 + F_n(2)m_2 + \cdots + F_n(s)m_s \equiv 0(\mathrm{mod}\ F_n)$$

*has no solution in the domain*

$$||\mathbf{m}|| \leqslant c(\eta)F_n^{\frac{1}{2}+\frac{1}{2^{s+1}\ln 2}+\frac{1}{2^{2s+3}}}, \quad \mathbf{m} \neq \mathbf{0}.$$

From Lemmas 7.5 and 7.6, and Theorem 7.9, we can derive

**Theorem 7.12** . *Suppose that* $\alpha > 1$. *Then*

$$\sup_{f\in E_s^\alpha(C)} \Big| \int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n}\sum_{k=1}^{n} f\Big(\frac{c_1k}{n}, \frac{c_2k}{n}, \cdots, \frac{c_sk}{n}\Big)\Big| \leqslant Cc(\mathscr{R}_s, \alpha, \varepsilon)n^{-\frac{\alpha}{2}-\frac{\alpha}{2(s-1)}+\varepsilon}, \quad s = \frac{\varphi(m)}{2},$$

and

$$\sup_{f\in E_s^\alpha(C)}\left|\int_{G_s} f(\mathbf{x})d\mathbf{x}-\frac{1}{F_n}\sum_{k=1}^{F_n} f\Big(\frac{k}{F_n},\frac{F_n(2)k}{F_n},\cdots,\frac{F_n(s)k}{F_n}\Big)\right| \leqslant Cc(\eta,\alpha)F_n^{-\frac{\alpha}{2}-\frac{\alpha}{2^{s+1}\ln 2}-\frac{\alpha}{2^{2s+4}}}. \tag{7.18}$$

Using Theorems 7.11 and 2.9 instead of Theorems 7.9 and 2.8 respectively, we have

**Theorem 7.13** *The right hand side of* (7.18) *can be replaced by* $Cc(\eta,\alpha)F_n^{-\alpha}\ln 3F_n$ for $s=2$ and $Cc(\eta,\alpha,\varepsilon)F_n^{-3/4+\varepsilon}$ for $s=3$ *respectively.*

we may also obtain the corresponding quadrature formula for the class of functions $Q_s^\alpha(C)$ by the use of Theorem 7.10 instead of Theorem 7.9.

## 7.6 The *glp* set and numerical integration

**Lemma 7.7** *Suppose that* $\alpha>1$ *and* $N\geqslant 1$. *Then*

$$\sum_{||\mathbf{m}||\leqslant N} 1\leqslant 3^s N(\ln 3N)^{s-1} \tag{7.19}$$

*and*

$$\sum_{||\mathbf{m}||\geqslant N}\frac{1}{||\mathbf{m}||^\alpha}\leqslant (5\zeta(\alpha))^s N^{-\alpha+1}(\ln 3N)^{s-1}. \tag{7.20}$$

*Proof.* (7.19) is obviously true for $s=1$. Now suppose that $k\geqslant 1$ and (7.19) holds for $s\leqslant k$. Then

$$\sum_{\bar m_1\cdots\bar m_{k+1}\leqslant N}1=\sum_{\bar m_1\leqslant N}\sum_{\bar m_2\cdots\bar m_{k+1}\leqslant N/\bar m_1}1\leqslant 3^kN(\ln 3N)^{k-1}\sum_{\bar m_1\leqslant N}\frac{1}{\bar m} < 3^{k+1}N(\ln 3N)^k.$$

Hence (7.19) follows by mathematical induction.

For $s=1$,

$$\sum_{\bar m\geqslant N}\frac{1}{\bar m^\alpha}=2\sum_{m\geqslant N}\frac{1}{m^\alpha}\leqslant\frac{2}{N^\alpha}+2\int_N^\infty\frac{dt}{t^\alpha} =\frac{2}{N^\alpha}+\frac{2}{(\alpha-1)N^{\alpha-1}}<5\zeta(\alpha)N^{-\alpha+1}.$$

Now suppose that $k \geqslant 1$ and (7.20) holds for $s \leqslant k$. Then

$$\begin{aligned}\sum_{\bar{m}_1\cdots\bar{m}_{k+1}\geqslant N} \frac{1}{(\bar{m}_1\cdots\bar{m}_{k+1})^\alpha} &= \sum_{\bar{m}_1\leqslant N}\frac{1}{\bar{m}_1^\alpha}\sum_{\bar{m}_2\cdots\bar{m}_{k+1}\geqslant N/\bar{m}_1}\frac{1}{(\bar{m}_2\cdots\bar{m}_{k+1})^\alpha}\\ &+\sum_{\bar{m}_1>N}\frac{1}{\bar{m}_1^\alpha}\sum_{\bar{m}_2\cdots\bar{m}_{k+1}\geqslant 1}\frac{1}{(\bar{m}_2\cdots\bar{m}_{k+1})^\alpha}\\ &< (5\zeta(\alpha))^k N^{-\alpha+1}(\ln 3N)^{k-1}\sum_{\bar{m}_1\leqslant N}\frac{1}{\bar{m}_1}+(3\zeta(\alpha))^k\sum_{\bar{m}_1>N}\frac{1}{\bar{m}_1^\alpha}\\ &< 3(5\zeta(\alpha))^k N^{-\alpha+1}(\ln 3N)^k+(3\zeta(\alpha))^{k+1}N^{-\alpha+1}\\ &< (5\zeta(\alpha))^{k+1}N^{-\alpha+1}(\ln 3N)^k.\end{aligned}$$

The lemma follows.

**Lemma 7.8** *Suppose that $0<\delta<1$. Then there exist no less than $p-[\delta p]$ integers in the interval $1\leqslant a\leqslant p$ such that the congruence*

$$(\mathbf{a},\mathbf{m}) = m_1+m_2a+\cdots+m_sa^{s-1}\equiv 0(\bmod p) \tag{7.21}$$

*has no solution in the domain*

$$||\mathbf{m}|| \leqslant \delta s^{-1}3^{-s}p(\ln 3p)^{-(s-1)},\quad \mathbf{m}\neq\mathbf{0}. \tag{7.22}$$

*Proof.* We may suppose that $\delta s^{-1}3^{-s}p(\ln 3p)^{-(s-1)}\geqslant 1$. Otherwise the lemma evidently holds. Since the number of solutions of the congruence (7.21) is at most $s-1$ in the interval $1\leqslant a\leqslant p$ for any given $\mathbf{m}$ belonging to (7.22) (Cf. Lemma 4.5), the total number of solutions of the congruence (7.21) with $\mathbf{m}$ satisfying (7.22) is at most

$$\begin{aligned}&\sum_{||\mathbf{m}||\leqslant\delta s^{-1}3^{-s}p(\ln 3p)^{-(s-1)}}\ \sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0(\bmod\ p)\\ 1\leqslant a\leqslant p}} 1\\ &\leqslant (s-1)3^s(\ln 3p)^{s-1}\delta s^{-1}3^{-s}p(\ln 3p)^{-(s-1)}<\delta p.\end{aligned}$$

Hence there exist at least $p-[\delta p]$ integers in the interval $1\leqslant a\leqslant p$ such that the congruence (7.21) has no solution in the domain (7.22).The lemma is proved.

**Theorem 7.14** *There exists an integral vector $\mathbf{a}=(a_1,\cdots,a_s)$ depending only on $p$ such that*

$$\begin{aligned}&\sup_{f\in E_s^\alpha(C)}\Big|\int_{G_s} f(\mathbf{x})d\mathbf{x}-\frac{1}{p}\sum_{k=1}^{p} f\Big(\frac{k\mathbf{a}}{p}\Big)\Big|\\ &\qquad< Cc(\alpha,s)p^{-\alpha}(\ln p)^{(\alpha+1)(s-1)},\quad \alpha>1\end{aligned}$$

*and*

$$\begin{aligned}&\sup_{f\in Q_s^\alpha(C)}\Big|\int_{G_s} f(\mathbf{x})d\mathbf{x}-\frac{1}{p}\sum_{k=1}^{p} f\Big(\frac{k\mathbf{a}}{p}\Big)\Big|\\ &\qquad< Cc(\alpha,s)p^{-\alpha}(\ln p)^{(\alpha+1)(s-1)},\quad 0<\alpha\leqslant 1.\end{aligned}$$

*Proof.* Let $M = 2^{-1}s^{-1}3^{-s}p(\ln 3p)^{-(s-1)}$. Let $A$ denote the set of integers in the interval $1 \leqslant a \leqslant p$ such that the congruence (7.21) has no solution in the domain

$$||\mathbf{m}|| \leqslant M. \quad \mathbf{m} \neq \mathbf{0}.$$

Then it follows from Lemma 7.8 that the number of elements of $A$ is no less than $\frac{p+1}{2}$. Let $a_1 = 1, a_2 = a, \cdots, a_s = a^{s-1}$, where $a \in A$. Then the theorem follows from Theorem 7.11.

We shall give another proof of Theorem 7.14 with a slight modification of error term in the following.

**Theorem 7.15** *Suppose that* $\alpha > 1$. *Then there exists an integral vector* $\mathbf{a}(=\mathbf{a}(\mathbf{p}))$ *such that*

$$\sup_{f \in E_s^\alpha(C)} \left| \int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{p}\sum_{k=1}^{p} f\left(\frac{k\mathbf{a}}{p}\right)\right| < Cc(\alpha, s)p^{-\alpha}(\ln p)^{\alpha(s-1)}$$

*Proof.* The notations introduced in Theorem 7.14 are also used here. Let $\mathbf{a} = (1, a, \cdots, a^{s-1})$ and

$$\Omega(a) = \sum_{(\mathbf{a},\mathbf{m})\equiv 0 (\mathrm{mod}\ p)} \frac{1}{||\mathbf{m}||^\alpha}.$$

Then

$$\sup_{f \in E_s^\alpha(C)} \left| \int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{p}\sum_{k=1}^{p} f\left(\frac{k\mathbf{a}}{p}\right)\right| \leqslant C\Omega(a) \tag{7.23}$$

and by Lemma 7.7,

$$\begin{aligned}
\sum_{a\in A} Q(a) &= \sum_{a \in A} \sum_{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ p)}{}^{'} \frac{1}{||\mathbf{m}||^\alpha} \\
&\leqslant \sum_{||\mathbf{m}|| \geqslant M} \sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ p) \\ 1 \leqslant \alpha \leqslant p}} \frac{1}{||\mathbf{m}||^\alpha} \\
&\leqslant p \sum{}' \frac{1}{||p\mathbf{m}||^\alpha} + (s-1) \sum_{||\mathbf{m}|| \geqslant M} \frac{1}{||\mathbf{m}||^\alpha} \\
&< s(2\zeta(\alpha)+1)^s p^{-\alpha+1} + s(5\zeta(\alpha))^s M^{-\alpha+1}(\ln 3p)^{s-1} \\
&< \frac{1}{3}(2s)^\alpha (3^\alpha 5\zeta(\alpha))^s p^{-\alpha+1}(\ln 3p)^{\alpha(s-1)}.
\end{aligned} \tag{7.24}$$

There are at most $[p/3]$ elements of $A$ such that the corresponding $Q(a)$ satisfies

$$\Omega(a) \geqslant (2s)^\alpha (3^\alpha 5\zeta(\alpha))^s p^{-\alpha}(\ln 3p)^{\alpha(s-1)}.$$

Otherwise we have

$$\sum_{a\in A}\Omega(a)\geqslant\left(\left[\frac{p}{3}\right]+1\right)(2s)^{\alpha}3^{\alpha s}(5\zeta(\alpha))^{s}p^{-\alpha}(\ln 3p)^{\alpha(s-1)}$$
$$>\frac{1}{3}(2s)^{\alpha}(3^{\alpha}5\zeta(\alpha))^{s}p^{-\alpha+1}(\ln 3p)^{\alpha(s-1)}$$

which leads to a contradiction with (7.24). Since

$$\frac{p+1}{2}-\left[\frac{p}{3}\right]\geqslant 1,$$

so there exists at least an integer $a$ of $A$ such that

$$\Omega(a)<(2s)^{\alpha}(3^{\alpha}5\zeta(\alpha))^{s}p^{-\alpha}(\ln 3p)^{\alpha(s-1)} \tag{7.25}$$

and the theorem follows from (7.23) and (7.25).

**Theorem 7.16** *Suppose that* $0<\alpha\leqslant 1$. *Then there exists and integral vector* $\mathbf{a}(=\mathbf{a}(p))$ *such that*

$$\sup_{f\in Q_s^{\alpha}(C)}\left|\int_{G_s}f(\mathbf{x})d\mathbf{x}-\frac{1}{p}\sum_{k=1}^{p}f\left(\frac{k\mathbf{a}}{p}\right)\right|$$
$$<Cc(\alpha,s,\varepsilon)p^{-\alpha}(\ln p)^{s-1+\varepsilon}.$$

*Proof.* By (7.7), we have

$$\sup_{f\in Q_s^{\alpha}(C)}\left|\int_{G_s}f(\mathbf{x})d\mathbf{x}-\frac{1}{p}\sum_{k=1}^{p}f\left(\frac{k\mathbf{a}}{p}\right)\right|$$
$$\leqslant Cc(\alpha,s)p^{-\alpha}(\ln p)^{s-1}+\Lambda(a),$$

where $\mathbf{a}=(1,a,\cdots,a^{s-1})$ and

$$\Lambda(a)=\sup_{f\in Q_s^{\alpha}(C)}\sum_{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ p)}{}'\ \sum_{t_0<\log_2 p}{}''\ |C_{\mathbf{t}}(\mathbf{m})|.$$

By Theorem 6.2,

$$\begin{aligned}\sum_{a\in A}\Lambda(a) &= \sup_{f\in Q_s^\alpha(C)}\sum_{a\in A}\sum_{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ p)}{}' \sum_{t_0<\log_2 p}{}'' |C_{\mathbf{t}}(\mathbf{m})| \\ &\leqslant \sup_{f\in Q_s^\alpha(C)}\sum_{M<||\mathbf{m}||<p}\sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ p)\\ 1\leqslant\alpha\leqslant p}} \sum{}'' |C_{\mathbf{t}}(\mathbf{m})| \\ &\leqslant \sup_{f\in Q_s^\alpha(C)}(s-1)\sum_{M<||\mathbf{m}||<p}|C(\mathbf{m})| \leqslant Cc(\alpha,s)\sum_{M<||\mathbf{m}||<p}\frac{1}{||\mathbf{m}||^\alpha} \\ &= Cc(\alpha,s)\sum_{M<||\mathbf{m}||<p}\frac{||\mathbf{m}||^{1-\alpha+\varepsilon/s}}{||\mathbf{m}||^{1+\varepsilon/s}} \\ &\leqslant Cc(\alpha,s)p^{1-\alpha+\varepsilon/s}\sum_{||\mathbf{m}||>M}\frac{1}{||\mathbf{m}||^{1+\varepsilon/s}}.\end{aligned}$$

Hence

$$\begin{aligned}\sum_{a\in A}\Lambda(a) &\leqslant Cc(\alpha,s,\varepsilon)p^{1-\alpha+\varepsilon/s}M^{-\varepsilon/s}(\ln M)^{s-1} \\ &\leqslant Cc(\alpha,s,\varepsilon)p^{1-\alpha}(\ln p)^{s-1+\varepsilon}\end{aligned}$$

by Lemma 7.7 and so there exists an integer $a$ of $A$ such that

$$\Lambda(a)\leqslant\frac{2}{p+1}\sum_{a\in A}\Lambda(a)\leqslant Cc(\alpha,s,\varepsilon)p^{-\alpha}(\ln p)^{s-1+\varepsilon}.$$

The theorem follows.

## 7.7 The Sarygin theorem

**Theorem 7.17** *For any given $n$ points of $G_s$*

$$P(k)=(x_1^{(k)},\cdots,x_s^{(k)}),\quad k=1,\cdots,n,$$

*there exists $f\in E_s^\alpha(C)$ such that*

$$f(P(k))=0,\quad k=1,\cdots,n$$

*and*

$$\int_{G_s}f(\mathbf{x})d\mathbf{x}\geqslant Cc(\alpha,s)n^{-\alpha}(\ln n)^{s-1}.$$

*Proof.* Let $t$ denote the integer satisfying $2^{t-1}\leqslant n<2^t$. Then we add any $2^t-n$ points

$$P(k)=(x_1^{(k)},\cdots,x_s^{(k)}),\quad k=n+1,\cdots,2^t$$

and consider the integral vectors

$$\mathbf{r} = (r_1, \cdots, r_s),$$

where $r_1 + \cdots + r_s = t, r_i \geqslant 0 (1 \leqslant i \leqslant s)$. Let $M(\mathbf{r})$ denote the set of integral vectors $\mathbf{m}$ such that

$$\bar{m}_i \leqslant 2^{r_i}, \quad i = 1, \cdots, s.$$

Then the number of elements of $M(\mathbf{r})$ is no less than $2^{t+1} + 1$. Now we shall prove that for any given $\mathbf{r}$, there exists a trigonometrical polynomial

$$T_r(\mathbf{x}) = \sum_{\mathbf{m} \in M(\mathbf{r})} C_{\mathbf{r}}(\mathbf{m}) e^{2\pi i (\mathbf{m}, \mathbf{x})} \not\equiv 0 \tag{7.26}$$

such that

$$T_{\mathbf{r}}(P(k)) = 0, \quad k = 1, \cdots, 2^t. \tag{7.27}$$

Indeed, since (7.27) is a system of linear equations of the Fourier coefficients $C_r(\mathbf{m})$ and the number of unknowns is greater then the number of equations, there exist $C_{\mathbf{r}}(\mathbf{m})$'s not all zero such that $T_{\mathbf{r}}(\mathbf{x})$ satisfies (7.26) and (7.27).

Let

$$T_{\mathbf{r}}^{(0)}(\mathbf{x}) = \frac{T_{\mathbf{r}}(\mathbf{x}) e^{-2\pi i (\mathbf{m}', \mathbf{x})}}{C_{\mathbf{r}}(\mathbf{m}') 2^{\alpha(t+s)}},$$

where $\mathbf{m}' = (m_1', \cdots, m_s')$ and

$$|C_{\mathbf{r}}(\mathbf{m}')| = \max_{\mathbf{m} \in M(\mathbf{r})} |C_{\mathbf{r}}(\mathbf{m})|.$$

Now we proceed to prove that we may choose constant $\chi = c(\alpha, s)$ such that

$$f(\mathbf{x}) = C\chi \sum_{\mathbf{r}} T_{\mathbf{r}}^{(0)}(\mathbf{x}) \in E_s^{\alpha}(C).$$

Since $e^{2\pi i (\mathbf{m}, \mathbf{x})}$ may appear only in those $T_{\mathbf{r}}^{(0)}(\mathbf{x})$ with

$$\bar{m}_1 \leqslant 2^{r_1+1}, \cdots, \quad \bar{m}_s \leqslant 2^{r_s+1},$$

so $r_1$ satisfies

$$\begin{aligned} \log_2 \bar{m}_1 - 1 \leqslant r_1 &= t - r_2 - \cdots - r_s \\ &\leqslant t - \log_2 \bar{m}_2 - \cdots - \log_2 \bar{m}_s + s - 1 \end{aligned}$$

and so the values that may be taken by $r_1$ and also the other $r_i'$s do not exceed

$$t - \log_2 \bar{m}_1 - \cdots - \log_2 \bar{m}_s + s + 1 = \log_2 \frac{2^t}{\|\mathbf{m}\|} + s + 1.$$

Consequently, the number of polynomials $T_{\mathbf{r}}^{(0)}(\mathbf{x})$ which contain the term $e^{2\pi i(\mathbf{m},\mathbf{x})}$ is at most

$$\left(\log_2 \frac{2^{t+s}}{||\mathbf{m}||} + 1\right)^s.$$

Take

$$\chi = \inf_{0<y\leqslant 1} y^{-\alpha}\left(\log_2 \frac{1}{y} + 1\right)^{-s}.$$

Then $\chi = c(\alpha, s)$ and

$$C\chi \frac{\left(\log_2 \frac{2^{t+s}}{||\mathbf{m}||} + 1\right)^s}{2^{\alpha(t+s)}} \leqslant C||\mathbf{m}||^{-\alpha}.$$

Hence $f \in E_s^\alpha(C)$.

By (7.27), we have

$$f(P(k)) = 0, \quad k = 1, 2, \cdots, 2^t$$

and by Lemma 7.1,

$$\begin{aligned} \int_{G_s} f(\mathbf{x})d\mathbf{x} &= C\chi 2^{-\alpha(t+s)} \sum_{\mathbf{r}} 1 = C\chi 2^{-\alpha(t+s)} C_{s-1}^{t+s-1} \\ &\geqslant Cc(\alpha, s)n^{-\alpha}(\ln n)^{s-1}. \end{aligned}$$

The theorem is proved.

We know by Theorem 7.17 that the error terms for the quadrature formulas given by Theorems 7.4 and 7.15 are of the best possible kind in principal order and the error term given by Theorem 7.13 for $s = 2$ is of the best possible kind apart from some possible improvement about the constant $c(\eta, \alpha)$.

## 7.8 The mean error of the quadrature formula

Let

$$c(s, \varepsilon) = 2((2\zeta(1+\varepsilon) + 1)^s - 1),$$

where $0 < \varepsilon < 1$. Let $\Omega(= \Omega(\varepsilon))$ denote the set of points $\boldsymbol{\gamma}$ of $G_s$ such that the inequality

$$\langle(\mathbf{m}, \boldsymbol{\gamma})\rangle \geqslant \varepsilon c(s, \varepsilon)^{-1}||\mathbf{m}||^{-1-\varepsilon}$$

holds for any integral vector $\mathbf{m} \neq \mathbf{0}$. Then by Theorem 4.10, the Lebesgue measure of $\Omega$ satisfies

$$\text{mes } \Omega > 1 - \varepsilon.$$

Let

$$S(n, \Omega, f) = \int_\Omega |S(n, \gamma, f)| d\gamma,$$

where $f \in Q_s^\alpha(C) \Big(\alpha > \frac{1}{2}\Big)$ and

$$S(n, \gamma, f) = \int_{G_s} f(\mathbf{x}) d\mathbf{x} - \frac{1}{(2n+1)^l} \sum_{k=-nl}^{nl} \mu_{n,l,k} f(k\gamma),$$

in which $l$ is the least integer$\geqslant \alpha + 1$ and $\mu_{n,l,k}$ is defined in §7.3. $S(n, \Omega, f)$ is called the mean error of the quadrature formula given by good points.

**Theorem 7.18** *Suppose that* $\alpha > \frac{1}{2}$. *Then*

$$\sup_{f \in Q_s^\alpha(C)} S(n, \Omega, f) \leqslant Cc(\alpha, s, \varepsilon) n^{-\alpha - \frac{1}{2} + \varepsilon}.$$

To proof Theorem 7.18, we shall need

**Lemma 7.9** *Suppose that* $\alpha > \frac{1}{2}$ *and* $f \in Q_s^\alpha(C)$. *Then*

$$\sum |C(\mathbf{m})|^2 ||\mathbf{m}||^{2(\alpha - \varepsilon)} \leqslant C^2 c(s, \varepsilon),$$

*where* $C(\mathbf{m})$ *denotes the Fourier coefficient of f.*

*Proof.* Since

$$\sum_{\mathbf{m}} |C_{\mathbf{t}}(\mathbf{m})|^2 = ||\varphi_{\mathbf{t}}||_2^2 \leqslant ||\varphi_{\mathbf{t}}||^2 \leqslant C^2 2^{-2\alpha t_0}$$

and $C_{\mathbf{t}}(\mathbf{m}) = 0$ for $||\mathbf{m}|| \geqslant 2^{t_0}$, therefore

$$\sum_{\mathbf{m}} |C_{\mathbf{t}}(\mathbf{m})|^2 ||\mathbf{m}||^{2(\alpha-\varepsilon)} \leqslant 2^{2(\alpha-\varepsilon)t_0} \sum_{\mathbf{m}} |C_{\mathbf{t}}(\mathbf{m})|^2 \leqslant C^2 2^{-2\varepsilon t_0}$$

and so

$$\begin{aligned}\sum_{\mathbf{m}} |C(\mathbf{m})|^2 ||\mathbf{m}||^{2(\alpha-\varepsilon)} &= \sum_{\mathbf{m}} \Big| {\sum_{\mathbf{t}}}'' C_{\mathbf{t}}(\mathbf{m}) \Big|^2 ||\mathbf{m}||^{2(\alpha-\varepsilon)} \\ &\leqslant \sum_{\mathbf{m}} \Big( {\sum_{\mathbf{t}}}'' |C_{\mathbf{t}}(\mathbf{m})|^2 2^{\varepsilon t_0} \Big) \Big( {\sum_{\mathbf{t}}}'' 2^{-\varepsilon t_0} \Big) ||\mathbf{m}||^{2(\alpha-\varepsilon)} \\ &\leqslant c(s, \varepsilon) {\sum_{\mathbf{t}}}'' 2^{\varepsilon t_0} \sum_{\mathbf{m}} |C_{\mathbf{t}}(\mathbf{m})|^2 ||\mathbf{m}||^{2(\alpha-\varepsilon)} \\ &\leqslant C^2 c(s, \varepsilon) {\sum_{\mathbf{t}}}'' 2^{-\varepsilon t_0} = C^2 c(s, \varepsilon).\end{aligned}$$

The lemma is proved.

The proof of Theorem 7.18. We may suppose that $\varepsilon < \alpha - \frac{1}{2}$. Then by the argument of §7.3,

$$|S(n,\gamma,f)| \leqslant \sum{}' \frac{|C(\mathbf{m})|}{(2n+1)^l}\Big|\sum_{k=-n}^{n} e^{2\pi i(\mathbf{m},\boldsymbol{\gamma})k}\Big|^l.$$

Hence it follows by Schwarz's inequality that

$$S(n,\Omega,f) \leqslant \int_{\Omega} S_1^{1/2}S_2^{1/2}d\gamma, \tag{7.28}$$

where

$$S_1 = \sum{}'|C(\mathbf{m})|^2\frac{||\mathbf{m}||^{2(\alpha-\varepsilon/2)}}{(2n+1)^2}\Big|\sum_{k=-n}^{n} e^{2\pi i(\mathbf{m},\boldsymbol{\gamma})k}\Big|^2$$

and

$$S_2 = \sum{}' \frac{1}{(2n+1)^{2(l-1)}||\mathbf{m}||^{2(\alpha-\varepsilon/2)}}\Big|\sum_{k=-n}^{n} e^{2\pi i(\mathbf{m},\boldsymbol{\gamma})k}\Big|^{2(l-1)}$$

By (6.9) and Lemma 7.9,

$$\begin{aligned}\int_{G_s} S_1 d\gamma &= \sum{}'|C(\mathbf{m})|^2\frac{||\mathbf{m}||^{2(\alpha-\varepsilon/2)}}{(2n+1)^2}\int_{G_s}\Big|\sum_{k=-n}^{n} e^{2\pi i(\mathbf{m},\boldsymbol{\gamma})k}\Big|^2 d\gamma\\ &= \sum{}'|C(\mathbf{m})|^2\frac{||\mathbf{m}||^{2(\alpha-\varepsilon/2)}}{(2n+1)^2}\int_{G_s}\Big(\sum_{k=-n}^{n} e^{2\pi i(\mathbf{m},\boldsymbol{\gamma})k}\Big)\Big(\sum_{j=-n}^{n} e^{-2\pi i(\mathbf{m},\boldsymbol{\gamma})j}\Big)d\gamma\\ &= \frac{1}{2n+1}\sum{}'|C(\mathbf{m})|^2||\mathbf{m}||^{2(\alpha-\varepsilon/2)} \leqslant C^2c(s,\varepsilon)n^{-1}.\end{aligned}$$

Since $2(l-1) \geqslant 2\alpha > 1$ and $2\alpha - \varepsilon > 1$, hence in a way similar to the proof of Theorem 7.4,

$$S_2 \leqslant c(\alpha,s,\varepsilon)n^{-2\alpha+2\varepsilon}$$

and so by (7.28) and Schwarz's inequality, we have

$$\begin{aligned}S(n,\Omega,f) &\leqslant c(\alpha,s,\varepsilon)n^{-\alpha+\varepsilon}\int_{G_s} S_1^{1/2}d\gamma\\ &\leqslant c(\alpha,s,\varepsilon)n^{-\alpha+\varepsilon}\Big(\int_{G_s} S_1 d\gamma\Big)^{1/2}\\ &\leqslant Cc(\alpha,s,\varepsilon)n^{-\alpha-\frac{1}{2}+\varepsilon}.\end{aligned}$$

The theorem is proved.

We know from Theorem 7.18 that the mean error of the quadrature formula given by the good points is better by a factor $O(n^{-\frac{1}{2}})$ compared with the error term of the

quadrature formula given by Theorem 7.4 for the class of functions $Q_s^\alpha(C)\left(\alpha > \frac{1}{2}\right)$. Hence we may expect that perhaps the error term of the quadrature formula given by good points is $O(n^{-\alpha-\frac{1}{2}+\varepsilon})$.

## 7.9 Continuation

Suppose that $0 < \varepsilon < \frac{1}{2}$. Then it follows by Lemma 7.8 that for any prime $p$ in the interval $n/2 < p \leqslant n$, there exist no less that $p - [\varepsilon p] \geqslant (1-\varepsilon)p$ integral vectors $\mathbf{a} = (1, a, \cdots, a^{s-1})(1 \leqslant a \leqslant p)$ such that the non-trivial solution $\mathbf{m}$ of the congruence

$$(\mathbf{a}, \mathbf{m}) \equiv 0 (\text{mod } p)$$

satisfy

$$||\mathbf{m}|| \geqslant \varepsilon s^{-1} 3^{-s} p (\ln 3p)^{-(s-1)} \geqslant \varepsilon 2^{-1} s^{-1} 3^{-s} n (\ln 3n)^{-(s-1)} = M (\text{say}).$$

Let $\omega (= \omega(\varepsilon, n))$ be the set of these $(\mathbf{a}, p)$. Denote the number of elements of $\omega$ by $|\omega|$. Then by the prime number theorem (Cf. Hua Loo Keng [2], Chap. 9),

$$|\omega| \geqslant \sum_{\frac{n}{2} < p \leqslant n} (1-\varepsilon)p \geqslant \frac{cn^2}{\ln n}. \tag{7.29}$$

Let

$$S(n, \omega, f) = \frac{1}{|\omega|} \sum_{(\mathbf{a}, p) \in \omega} |S(p, \mathbf{a}, f)|,$$

where $f \in Q_s^\alpha(C)\left(\alpha > \frac{1}{2}\right)$ and

$$S(p, \mathbf{a}, f) = \int_{G_s} f(x) dx - \frac{1}{p} \sum_{k=1}^{p} f\left(\frac{k\mathbf{a}}{p}\right).$$

$S(n, \omega, f)$ is called the mean error of the quadrature formula given by good lattice points.

**Theorem 7.19** . *Suppose that* $\alpha > \frac{1}{2}$. *Then*

$$\sup_{f \in Q_s^\alpha(C)} S(n, \omega, f) \leqslant Cc(\alpha, s, \varepsilon) n^{-\alpha - \frac{1}{2} + \varepsilon}.$$

To prove Theorem 7.19, we shall need

**Lemma 7.10**

$$A(\omega, \mathbf{m}) = \sum_{\substack{(\mathbf{a},p)\in\omega \\ (\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ p)}} 1 \leqslant n \log_{\frac{n}{2}}(s\|\mathbf{m}\|n^{s-1}).$$

*Proof.* Let $\sigma_m(l)$ be the number of prime divisors of $l$ which are $\geqslant m$. Then

$$\sigma_m(l) \leqslant \log_m l.$$

Hence

$$\begin{aligned} A(\omega, \mathbf{m}) &\leqslant \sum_{1\leqslant a\leqslant p} \sum_{\substack{n/2<p\leqslant n \\ (\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ p)}} 1 \leqslant \sum_{1\leqslant a\leqslant p} \sigma_{\frac{n}{2}}(\mathbf{a}, \mathbf{m}) \\ &\leqslant n \log_{\frac{n}{2}}(s\|\mathbf{m}\|n^{s-1}). \end{aligned}$$

The lemma is proved.

The proof of Theorem 7.19. We may suppose that $\varepsilon < \alpha - \dfrac{1}{2}$. Then by the argument of §7.5,

$$|S(p, \mathbf{a}, f)| = \Big| \sum_{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ p)}' C(\mathbf{m}) \Big|.$$

Hence by (7.29), Lemma 7.10 and Schwarz's inequality,

$$\begin{aligned} S(n, \omega, f) &\leqslant \frac{1}{|\omega|} \sum_{(\mathbf{a},p)\in\omega} \sum_{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ p)}' |C(\mathbf{m})| \\ &\leqslant \frac{1}{|\omega|} \sum_{\|\mathbf{m}\|\geqslant M} A(\omega, \mathbf{m})|C(\mathbf{m})| \\ &\leqslant \frac{1}{|\omega|} \Big(\sum |C(\mathbf{m})|^2 \|\mathbf{m}\|^{2(\alpha-\varepsilon/2)}\Big)^{1/2} \Big( \sum_{\|\mathbf{m}\|\geqslant M} A(\omega, \mathbf{m})^2 \|\mathbf{m}\|^{-2(\alpha-\varepsilon/2)} \Big)^{1/2} \\ &\leqslant Cc(s, \varepsilon) n^{-2}(\ln\, n) \Big( \sum_{\|\mathbf{m}\|\geqslant M} n^2(1 + \log_{\frac{n}{2}} \|\mathbf{m}\|)^2 \|\mathbf{m}\|^{-2(\alpha-\varepsilon/2)} \Big)^{1/2} \\ &\leqslant Cc(s, \varepsilon) n^{-1}(\ln\, n) \Big( \sum_{\|\mathbf{m}\|\geqslant M} \frac{\|\mathbf{m}\|^{-2\alpha+1+3\varepsilon/2}(\ln\|\mathbf{m}\|)^2}{\|\mathbf{m}\|^{1+\varepsilon/2}} \Big)^{1/2} \\ &\leqslant Cc(s, \varepsilon) n^{-1}(\ln\, n) M^{-\alpha+\frac{1}{2}+\frac{3}{4}\varepsilon} \Big( \sum \frac{(\ln\|\mathbf{m}\|)^2}{\|\mathbf{m}\|^{1+\varepsilon/2}} \Big)^{1/2} \\ &\leqslant Cc(s, \varepsilon) n^{-\alpha-\frac{1}{2}+\varepsilon}. \end{aligned}$$

The theorem is proved.

We know form Theorem 7.19 that the mean error of the quadrature formula given by the good lattice points in better by a factor $O(n^{-1/2})$ compared with the error

term of the quadrature formula given by Theorem 7.14 for the class of functions $Q_s^\alpha(C)\left(\alpha > \frac{1}{2}\right)$. Hence we may expect that perhaps the error term of the quadrature formula given by good lattice point is $O(n^{-\alpha-\frac{1}{2}+\varepsilon})$.

## Notes

N. M. Korobov [1] was the first to have proved a result with the same precision of Theorem 7.2 for the $p$ set $\left(\left\{\frac{k}{p^2}\right\}, \cdots, \left\{\frac{k^s}{p^2}\right\}\right)(1 \leqslant k \leqslant p^2)$ and Theorem 7.2 was proved by Hua Loo Keng and Wang Yuan [3]. Theorem 7.3 was proved by Yu. N. Sahov [4] (Cf. also V. M. Solodov [1]) for the case $\alpha > \frac{1}{2}$ and by Hua Loo Keng and Wang Yuan [6,7] for $\frac{1}{2} > \alpha \geqslant 0$.

R. D. Richtmyer [1] and L. Peck [1] proposed using the set $(\{r_1 k\}, \cdots, \{r_s k\})(k = 1, 2, \cdots)$ to evaluate the multiple integral. Theorem 7.4 was first proved by N. S. Bahvalov [1] and C. B. Haselgrove [1] (Cf. also Wang Yuan [2], N. M. Korobov [7], Hua Loo Keng and Wang Yuan [3,6,7], H. Niederreiter [2] and Wang Yuan [5]).

Theorem 7.8: Cf. N. M. Gelfand, A. S. Frolov and N. N. Cencov [1] and N. M. Korobov [7].

Theorem 7.11 was first proved by N. B. Bahvalov [1] (Cf. also Hua Loo Keng and Wang Yuan [3,6,7]). The case $s = 2$ of Theorem 7.13 was proved independently by Hua Loo Keng and Wang Yuan [1] and N. S. Bahvalov [1].

Theorem 4.15 was first proved by N. M. Korobov [2] with the error term $O(p^{-\alpha}(\ln p)^{\alpha s})$ and improved by N. S. Bahvalov [1] to $O(p^{-\alpha}(\ln p)^{-\alpha(s-1)})$ (Cf. also Wang Yuan [2]).

Theorem 7.17:Cf. I. F. Sarygin [2].

Theorems 7.18 and 7.19: Cf. N. B. Bahvalov [1,2] and Wang Yuan [2].

# Chapter 8

# Numerical Error for Quadrature Formula

## 8.1 The numerical error

We introduce the notations

$$S(n,\gamma,f)=\int_{G_s}f(\mathbf{x})d\mathbf{x}-\frac{1}{n}\sum_{k=-(n-1)}^{n-1}\left(1-\frac{|k|}{n}\right)f(k\gamma)$$

and

$$S(n,\mathbf{h},f)=\int_{G_s}f(\mathbf{x})d\mathbf{x}-\frac{1}{n}\sum_{k=1}^{n}f\left(\frac{k\mathbf{h}}{n}\right),$$

where $\gamma$ and $h$ denote real and integral vectors respectively.

**Theorem 8.1**

$$\sup_{f\in E_s^2(C)}|S(n,\gamma,f)|\leqslant CW_2(n,\gamma),$$

*where*

$$W_2(n,\gamma)=\frac{1}{n}\left(1+\frac{\pi^2}{3}\right)^s+\frac{2}{n}\sum_{k=1}^{n}\left(1-\frac{k}{n}\right)\prod_{v=1}^{s}(1+2\pi^2B_2(\{k\gamma_v\}))-1,$$

in which $B_2(x)=x^2-x+\dfrac{1}{6}$ denotes the Bernoulli polynomial (Cf. §6.5).

**Theorem 8.2**

$$\sup_{f\in E_s^2(C)}|S(n,\mathbf{h},f)|\leqslant CW_2(n,\mathbf{h}),$$

*where*

$$W_2(n,\mathbf{h})=\begin{cases}\dfrac{1}{n}\left(1+\dfrac{\pi^2}{3}\right)^s+\dfrac{2}{n}\displaystyle\sum_{k=1}^{\frac{n-1}{2}}\prod_{v=1}^{s}\left(1+2\pi^2B_2\left(\left\{\frac{kh_v}{n}\right\}\right)\right), & \text{if } 2\nmid n,\\ \dfrac{1}{n}\left(1+\dfrac{\pi^2}{3}\right)^s+\dfrac{1}{n}\left(1-\dfrac{\pi^2}{6}\right)^\mu\left(1+\dfrac{\pi^2}{3}\right)^{s-\mu} & \\ +\dfrac{2}{n}\displaystyle\sum_{k=1}^{\frac{n}{2}-1}\prod_{v=1}^{s}\left(1+2\pi^2B_2\left(\left\{\frac{kh_v}{n}\right\}\right)\right)-1, & \text{if } 2|n,\end{cases}$$

*in which $\mu$ denotes the number of odd integers of $h_v (1 \leqslant v \leqslant s)$.*

**Lemma 8.1**

$$\sum_{m=-\infty}^{\infty} \frac{e^{2\pi i m x}}{\bar{m}^2} = 1 + 2\pi^2 B_2(\{x\}).$$

*Proof.* Since

$$\begin{aligned}&\int_0^1 (1 + 2\pi^2 B_2(x)) e^{-2\pi i m x} dx \\ =& \int_0^1 \left(1 - \frac{\pi^2}{6} + \frac{\pi^2}{2}(1-2x)^2\right) e^{-2\pi i m x} dx = \bar{m}^{-2},\end{aligned}$$

the lemma follows.

The proof of Theorem 8.1. By Lemma 7.4,

$$\begin{aligned}W_2(n, \boldsymbol{\gamma}) &= n^{-1} \Sigma' \frac{1}{\|\mathbf{m}\|^2} \left| \sum_{k=-(n-1)}^{n-1} (n - |k|) e^{2\pi i (\mathbf{m}, \boldsymbol{\gamma}) k} \right| \\ &= n^{-2} \Sigma' \frac{1}{\|\mathbf{m}\|^2} \sum_{j=0}^{n-1} \sum_{k=-j}^{j} e^{2\pi i (\mathbf{m}, \boldsymbol{\gamma}) k}.\end{aligned}$$

Hence by Lemma 8.1

$$\begin{aligned}W_2(n, \gamma) &= n^{-2} \sum_{j=0}^{n-1} \sum_{k=-j}^{j} \left( \Sigma \frac{1}{\|\mathbf{m}\|^2} e^{2\pi i (\mathbf{m}, \gamma) k} - 1 \right) \\ &= n^{-2} \sum_{j=0}^{n-1} \sum_{k=-j}^{j} \prod_{v=1}^{s} \left( \sum_{m_v=-\infty}^{\infty} \frac{e^{2\pi i m_v \gamma_v k}}{\bar{m}v^2} \right) - 1 \\ &= n^{-2} \sum_{j=0}^{n-1} \sum_{k=-j}^{j} \prod_{v=1}^{s} (1 + 2\pi^2 B_2(\{k\gamma_v\})) - 1 \\ &= \frac{1}{n} \sum_{k=-n}^{n} \left( \frac{n - |k|}{n} \right) \prod_{v=1}^{s} (1 + 2\pi^2 B_2(\{k\gamma_v\})) - 1.\end{aligned}$$

Since $B_2(\{x\})$ is an even function of $x$, the theorem follows.

The proof of Theorem 8.2. Since

$$\begin{aligned}W_2(n, \mathbf{h}) &= \frac{1}{n} \sum_{k=1}^{n} \sum{}' \frac{e^{2\pi i (\boldsymbol{h}, \mathbf{m}) k / m}}{\|\mathbf{m}\|^2} \\ &= \frac{1}{n} \sum_{k=1}^{n} \prod_{v=1}^{s} \left( \sum_{m_v=-\infty}^{\infty} \frac{e^{2\pi i h_v m_v k / n}}{\bar{m}_v^2} \right) - 1\end{aligned}$$

$$= \frac{1}{n}\sum_{k=1}^{n}\prod_{v=1}^{s}\left(1+2\pi^2 B_2\left(\left\{\frac{kh_v}{n}\right\}\right)\right)-1.$$

The theorem follows.

Let

$$\sup_{f\in E_s^4(C)}|S(n,\mathbf{h},f)| \leqslant CW_4(n,\mathbf{h}).$$

Then in a way similar to Theorem 8.2, we may derive

**Theorem 8.3**

$$W_4(n,\mathbf{h}) = \begin{cases} \frac{1}{n}\left(1+\frac{\pi^4}{45}\right)^s + \frac{2}{n}\sum_{k=1}^{\frac{n-1}{2}}\prod_{v=1}^{s}\left(1-\frac{2\pi^4}{3}B_4\left(\left\{\frac{kh_v}{n}\right\}\right)\right)-1, \\ \qquad\qquad if\ 2\nmid n, \\ \frac{1}{n}\left(1+\frac{\pi^4}{45}\right)^s + \frac{1}{n}\left(1-\frac{7\pi^4}{360}\right)^{\mu}\left(1+\frac{\pi^4}{45}\right)^{s-\mu} \\ +\frac{2}{n}\sum_{k=1}^{\frac{n}{2}-1}\prod_{v=1}^{s}\left(1-\frac{2\pi^4}{3}B_4\left(\left\{\frac{kh_v}{n}\right\}\right)\right)-1, \quad if\ 2|n, \end{cases}$$

*where $\mu$ denotes the number of odd integers of $h_v (1 \leqslant v \leqslant s)$.*

We may also obtain the expressions for the

$$\sup_{f\in E_s^{2l}(C)}|S(n,\mathbf{h},f)|, \quad l=3,4,\cdots.$$

## 8.2 The comparison of good points

We shall construct the vectors $\gamma$ by the use of the cyclotomic field, the Dirichlet field and Theorem 4.13 and then we shall compare their corresponding $W_2(n,\gamma)$ as follows.

**Table 1** $s=3$

| $n$ | $W_2(n,(e,e^2,e^3))$ | $W_2\left(n,\left(\frac{\sqrt{5}-1}{2},\sqrt{2},\sqrt{10}\right)\right)$ |
|---|---|---|
| $10^2$ | $3.7687\times10^{-1}$ | $3.1740\times10^{-1}$ |
| $5\times10^2$ | $3.4290\times10^{-2}$ | $5.3640\times10^{-2}$ |
| $10^3$ | $1.1180\times10^{-2}$ | $2.2850\times10^{-2}$ |

**Table 2** $s=4$

| $n$ | $W_2\left(n,\left(2\cos\frac{2\pi}{11},\cdots,2\cos\frac{8\pi}{11}\right)\right)$ | $W_2(n,(e,\cdots,e^4))$ |
|---|---|---|
| $10^3$ | $1.5683\times10^{-1}$ | $6.4835\times10^{-1}$ |
| $1.5\times10^3$ | $1.0070\times10^{-1}$ | $5.5341\times10^{-1}$ |
| $3\times10^3$ | $4.0200\times10^{-2}$ | $3.7821\times10^{-1}$ |

**Table 3** $s = 5$

| $n$ | $W_2\left(n, \left(2\cos\frac{2\pi}{13}, \cdots, 2\cos\frac{10\pi}{13}\right)\right)$ | $W_2(n, (e, \cdots, e^5))$ |
|---|---|---|
| $10^4$ | $6.7819 \times 10^{-2}$ | $7.4518 \times 10^{-2}$ |
| $10^5$ | $3.6400 \times 10^{-3}$ | $3.6966 \times 10^{-3}$ |

**Table 4** $s = 7$

| $n$ | $W_2\left(n, \left(2\cos\frac{2\pi}{17}, \cdots, 2\cos\frac{14\pi}{17}\right)\right)$ | $W_2(n, (e, \cdots, e^7))$ | $W_2\left(n, \left(\frac{\sqrt{5}-1}{2}, \sqrt{2}, \cdots, \sqrt{30}\right)\right)$ |
|---|---|---|---|
| $10^3$ | $2.6730 \times 10$ | $3.1621 \times 10$ | $2.8308 \times 10$ |
| $10^4$ | $2.1492$ | $2.2513$ | $3.9925$ |
| $10^5$ | $1.9503 \times 10^{-1}$ | $1.9592 \times 10^{-1}$ | $3.0794 \times 10^{-1}$ |

*Remark.* It is suggested by the above tables that the $\gamma$ given by $\mathscr{R}_s$ or $\gamma = (e, e^2, \cdots, e^s)$ is more advantageous compared to the $\gamma$ given by $\mathscr{D}_s$ (Cf. P. J. Davis and P. Robinowitz [1]).

## 8.3 The computation of the $\eta$ set

Let $F_n (\equiv F_{s,n})$ be the generalized Fibonacci sequence of dimension s, i.e., the sequence of integers defined by the recurrent formula

$$F_0 = F_1 = \cdots = F_{s-2} = 0, \quad F_{s-1} = 1,$$
$$F_{n+s} = F_{n+s-1} + \cdots + F_{n+1} + F_n, \quad n \geqslant 0$$

(Cf. §2.8). Let $n = F_m, h_1 = 1, h_2 = F_{m+1} - F_m, \cdots, h_s = F_{m+s-1} - F_{m+s-2} - \cdots - F_{m+1} - F_m$. Then we have the $\eta$ set (Cf. §4.7) and the following examples.

1. $F_{2,m} = 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377$ for $0 \leqslant m \leqslant 15$ and

$$W_2(55, (1, 34)) \leqslant 3.8148 \times 10^{-2}.$$

2. Take $n$ to be $F_{13} = 401, F_{16} = 2,872$ and $F_{18} = 10,671$ for $s = 4$. Then we have

**Table 1** $(h_1 = 1)$

| $n$ | $h_2$ | $h_3$ | $h_4$ | $W_2(n, \mathbf{h})$ |
|---|---|---|---|---|
| 401 | 372 | 316 | 208 | $4.5260 \times 10^{-1}$ |
| 2,872 | 2,664 | 2,263 | 1,490 | $2.3845 \times 10^{-2}$ |
| 10,671 | 9,898 | 8,408 | 5,536 | $3.8520 \times 10^{-3}$ |

3. Take $n = F_{19} = 13,624, h_1 = 1, h_2 = 13,160, h_3 = 12,248, h_4 = 10,455, h_5 = 6,930$ for $s = 5$. Then

$$W_2(n, \mathbf{h}) \leqslant 3.0738 \times 10^{-2}.$$

4. Take $n = F_{21} = 29,970, h_1 = 1, h_2 = 29,478, h_3 = 28,502, h_4 = 26,566, h_5 = 22,726, h_6 = 15,109$ for $s = 6$. Then

$$W_2(n, \mathbf{h}) \leqslant 1.2002 \times 10^{-1}.$$

We may also obtain some data for $s \geqslant 7$ but the precision of the corresponding $W_2(n, \mathbf{h})$ is not so good (Cf. §8.7)

For $s = 3$, we suggest using the sequence of integers defined by the recurrent formula

$$G_0 = G_1 = 0, G_2 = 1, G_{m+3} = G_{m+1} + G_m, \quad m \geqslant 0$$

(Cf. §1.7). Let $n = G_m, h_1 = 1, h_2 = G_{m+1} - G_m, h_3 = G_{m+2} - G_m$.
Then we have

**Table 2** $(h_1 = 1)$

| $n$ | $h_2$ | $h_3$ | $W_2(n, \mathbf{h})$ |
|---|---|---|---|
| 151 | 49 | 114 | $1.5035 \times 10^{-1}$ |
| 1,081 | 351 | 816 | $1.1229 \times 10^{-2}$ |
| 1,897 | 616 | 1,432 | $3.8875 \times 10^{-3}$ |
| 13,581 | 4,410 | 10,252 | $1.2459 \times 10^{-4}$ |

## 8.4 The computation of the $\mathscr{R}_s$ set

Let $s = \dfrac{p-1}{2}$. Then

$$\rho_l = \frac{\sin \frac{\pi}{p} g^{l+1}}{\sin \frac{\pi}{p} g^l}, \quad 1 \leqslant l \leqslant s-1$$

is a set of independent units of $\mathscr{R}_s = Q\left(\cos \dfrac{2\pi}{p}\right)$, where $g$ denotes a primitive root mod $p$. In particular, if $g = 2$, then

$$\rho_l = 2\cos\frac{2\pi}{p}2^{l-1}, \quad 1 \leqslant l \leqslant s-1.$$

$$\xi^{(i)} = \rho_1^{(i)x_1} \cdots \rho_{s-1}^{(i)x_{s-1}}, \quad 2 \leqslant i \leqslant s,$$

where $\rho_j^{(i)} (2 \leqslant i \leqslant s)$ denote the conjugates of $\rho j$. Solving the system of linear equations

$$\ln|\xi^{(2)}| = \cdots = \ln|\xi^{(s)}|, \tag{8.1}$$

we obtain a unique set of ratios

$$\frac{x_1}{x_{s-1}}, \cdots, \frac{x_{s-2}}{x_{s-1}}.$$

Let $l$ be a given real number and $l_i$ and $l_{s-1}$ be the integers nearest to $lx_i/l_{s-1}(1 \leqslant i \leqslant s-2)$ and $l$ respectively. Then

$$\eta(=\eta_l) = |\rho_1^{l_1} \cdots \rho_{s-1}^{l_{s-1}}|$$

is an algebraic integer and so

$$n(=n_l) = \eta + \eta^{(2)} + \cdots + \eta^{(s)}$$

is a rational integer. For practical use, we may take $n$ to be the integer nearest to $\eta$ which is denoted by $n \doteq \eta$. Set

$$h_1 = 1, \quad h_i = n\left|\left\{2\cos\frac{2\pi i}{p}\right\}\right|, \quad 2 \leqslant i \leqslant s.$$

Then we have a $\mathscr{R}_s$ set (Cf. §4.6) and the following examples.

1. Take $\mathscr{R}_2 = Q\left(\cos\frac{2\pi}{5}\right)$. Then $n = F_{m+1}, h_1 = 1, h_2 = F_m$ is a $\eta$ set where $F_m = F_{2,m}$ (Cf. example 1 of §8.3).

2. Take $\mathscr{R}_3 = Q\left(\cos\frac{2\pi}{7}\right)$ and units

$$\varepsilon_1 = 2\cos\frac{6\pi}{7} = -1.8019\cdots,$$
$$\varepsilon_2 = 2\cos\frac{2\pi}{7} = 1.2473\cdots,$$
$$\varepsilon_3 = 2\cos\frac{4\pi}{7} = -0.4447\cdots.$$

Any two among these three units form a set of independent units of $\mathscr{R}_3$. Solving the equation

$$|\varepsilon_2^\alpha \varepsilon_2^\beta| = |\varepsilon_3^\alpha \varepsilon_1^\beta|,$$

we have

$$\frac{\alpha}{\beta} = 1.357\cdots \simeq \frac{4}{3},$$

where "$a \simeq b$" means that $a$ and $b$ are "approximately equal". Hence we have, for example

$$n = 418 \doteq \eta = \varepsilon_1^8\varepsilon_2^6, \quad h_1 = 1,$$
$$h_2 = 335 \doteq n\left\{\left|2\cos\frac{6\pi}{7}\right|\right\},$$
$$h_3 = 103 \doteq n\left\{\left|2\cos\frac{2\pi}{7}\right|\right\}$$

and also the following data:

| $n$ | $h_2$ | $h_3$ | $W_2(n, \mathbf{h})$ |
|---|---|---|---|
| 20 | 17 | 6 | 2.66 |
| 83 | 66 | 20 | $5.52 \times 10^{-1}$ |
| 418 | 335 | 103 | $3.71 \times 10^{-2}$ |
| 1,692 | 1,357 | 418 | $4.88 \times 10^{-3}$ |

3. Take $\mathscr{R}_5 = Q\left(\cos\frac{2\pi}{11}\right)$ and units

$$\varepsilon_1 = 2\cos\frac{10\pi}{11}, \quad \varepsilon_2 = 2\cos\frac{2\pi}{11}, \quad \varepsilon_3 = 2\cos\frac{8\pi}{11}, \quad \varepsilon_4 = 2\cos\frac{4\pi}{11}, \quad \varepsilon_5 = 2\cos\frac{6\pi}{11}.$$

Solving

$$|\varepsilon_2^\alpha \varepsilon_4^\beta \varepsilon_5^\gamma \varepsilon_3^\beta| = |\varepsilon_3^\alpha \varepsilon_5^\beta \varepsilon_2^\gamma \varepsilon_1^\delta| = |\varepsilon_4^\alpha \varepsilon_3^\beta \varepsilon_1^\gamma \varepsilon_5^\delta| = |\varepsilon_5^\alpha \varepsilon_1^\beta \varepsilon_4^\gamma \varepsilon_2^\delta|,$$

we have

$$\frac{\alpha}{\delta} = 1.412\cdots \simeq \frac{7}{5}, \quad \frac{\beta}{\delta} = 1.584\cdots \simeq \frac{8}{5}, \quad \frac{\gamma}{\delta} = 0.944\cdots \simeq \frac{5}{5}$$

and so

$$\begin{aligned} n &= 9,389 \doteq \eta = |\varepsilon_1^7 \varepsilon_2^8 \varepsilon_3^5 \varepsilon_4^5|, \quad h_1 = 1, \\ h_2 &= 8,628 \doteq n\left\{\left|2\cos\frac{10\pi}{11}\right|\right\}, \\ h_3 &= 6,408 \doteq n\left\{\left|2\cos\frac{2\pi}{11}\right|\right\}, \\ h_4 &= 2,908 \doteq n\left\{\left|2\cos\frac{8\pi}{11}\right|\right\}, \\ h_5 &= 7,800 \doteq n\left\{\left|2\cos\frac{4\pi}{11}\right|\right\}, \\ W_2&(h, \mathbf{h}) \leqslant 6.69 \times 10^{-2}. \end{aligned}$$

*Remark.* Suppose that $(n, \mathbf{h}(n))$ is an $\mathscr{R}_s$ set. Then we may obtain an $\mathscr{R}_{s'}$ set $(n, \mathbf{h}^*(n))$, where $\mathbf{h}^*(n)$ is obtained from the vector $\mathbf{h}(n)$ by neglecting any of its $s - s'$ components. The precision of $W_2(n, \mathbf{h}^*)$ is still good if $s/s'$ is close to 1. For example, suppose that $n = 462,891, h_1 = 1, h_2 = 450,265, h_3 = 412,730, h_4 = 351,310, h_5 = 267,681, h_6 = 164,124, h_7 = 43,464, h_8 = 371,882, h_9 = 277,266$. Then we obtain an $\mathscr{R}_7$ set $(n, \mathbf{h}_7^*)$ and an $\mathscr{R}_8$ set $(n, \mathbf{h}_8^*)$, where $\mathbf{h}_7^*$ and $\mathbf{h}_8^*$ are obtained rom $\mathbf{h}$ by neglecting $h_8$ and $h_9$ respectively. We also obtain

$$W_2(n, \mathbf{h}_7^*) \leqslant 1.9397 \times 10^{-2}$$

and

$$W_2(n, \mathbf{h}_8^*) \leqslant 1.6240 \times 10^{-1}.$$

Hence we may obtain the $\mathscr{R}_s$ set of any dimension $s$, although the cyclotomic fields used here are confined to the fields with dimension $\frac{p-1}{2}$.

## 8.5 Examples of other $\mathscr{F}_s$ sets

1. Take $\mathscr{D}_4 = Q(\sqrt{5}, \sqrt{2})$ and the units

$$\varepsilon_1 = \frac{1+\sqrt{5}}{2}, \quad \varepsilon_2 = 1+\sqrt{2}, \quad \varepsilon_3 = 3+\sqrt{10}.$$

Then from the unit

$$\varepsilon_1^6 \varepsilon_2^4 \varepsilon_3^2,$$

this gives

$$n = 11,574, \quad h_1 = 1, \quad h_2 = 7,153, \quad h_3 = 4,794, \quad h_4 = 1,878$$

and

$$W_2(n, \mathbf{h}) = 8.81 \times 10^{-3}.$$

2. Take $Q(\omega)$ and units

$$\varepsilon_1 = 2+\omega^2, \quad \varepsilon_2 = 3+2\omega,$$

where $\omega = \sqrt[4]{5}$ (Cf. L. Bernstein [1]). Solving the equation

$$\ln|2+\omega^2|^{x_1}|3-2\omega|^{x_2} = \ln|2-\omega^2|^{x_1}|3-2\omega i|^{x_2},$$

we have

$$\frac{x_1}{x_2} = 2.1200 \cdots.$$

From the unit

$$\varepsilon_1^4 \varepsilon_2^2,$$

we have

$$n = 2,889, \quad h_1 = 1, \quad h_2 = 1,431, \quad h_3 = 862, \quad h_4 = 993$$

and

$$W_2(n, \mathbf{h}) = 2.8626 \times 10^{-2}.$$

And from the unit

$$\varepsilon_1^6 \varepsilon_2^3,$$

we have

$$n = 310,563, \quad h_1 = 1, \quad h_2 = 153,837, \quad h_3 = 73,314, \quad h_4 = 106,741$$

and

$$W_2(n, \mathbf{h}) \leqslant 2.0948 \times 10^{-4}.$$

## 8.6 The computation of a *glp* set

We introduce the notations

$$H_1(z) = \frac{3^s}{p_1}\left(1 + 2\sum_{k=1}^{\frac{p_1-1}{2}}\prod_{v=0}^{s-1}\left(1 - 2\left\{\frac{kz^v}{p_1}\right\}\right)^2\right),$$

$$H_2(z) = \frac{3^s}{p_1p_2}\left(1 + 2\sum_{k=1}^{\frac{p_1p_2-1}{2}}\prod_{v=0}^{s-1}\left(1 - 2\left\{\frac{k(p_2b_1^v + p_1b_2^v)}{p_1p_2}\right\}\right)^2\right)$$

$$\cdots\cdots$$

$$H_t(z) = \frac{3^s}{q}\left(1 + 2\sum_{k=1}^{\frac{q-1}{2}}\prod_{v=0}^{s-1}\left(1 - 2\left\{\frac{k(q_1b_1^v + \cdots + q_{t-1}b_{t-1}^v + q_tz^v)}{q}\right\}\right)^2\right),$$

where $p_1, \cdots, p_t$ denote the different primes, $q = p_1 \cdots p_t, q_i = q/p_i (1 \leqslant i \leqslant t)$ and where $b_i$ denotes the integer such that $H_i(z)$ takes the minimum for $z = 1, \cdots, \dfrac{p_i - 1}{2} (1 \leqslant i \leqslant t)$.

**Lemma 8.2** *Let $\alpha > 1$. Let $q$ be a positive integer and $\mathbf{a} = (a_1, \cdots, a_s)$ be an integral vector. If $(a_i, q) = 1 (1 \leqslant i \leqslant s)$, then*

$$\sum_{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\, q)}{}' \frac{1}{\|\mathbf{m}\|^\alpha} - \sum_{\substack{(\mathbf{a},\mathbf{m}^{(0)})\equiv 0(\mathrm{mod}\, q)\\ -\frac{q}{2}<m_i^{(0)}\leqslant\frac{q}{2}}} \frac{1}{\|\mathbf{m}^{(0)}\|^\alpha} \leqslant s2^\alpha(2\zeta(\alpha)+1)^s q^{-\alpha}.$$

*Proof.* Since

$$\sum_{-\frac{q}{2}<m\leqslant\frac{q}{2}}\sum_{l=-\infty}^{\infty}\frac{1}{\overline{(m+lq)}^\alpha} \leqslant 1 + 2\sum_{n=1}^{\infty}\frac{1}{n^\alpha} = 1 + 2\zeta(\alpha)$$

and

$$\sum_{l=-\infty}^{\infty}{}' \frac{1}{\overline{(m+lq)}^\alpha} \leqslant 2\sum_{l=1}^{\infty}\frac{1}{\overline{q\left(1-\frac{1}{2}\right)}^\alpha} \leqslant 2\zeta(\alpha)\left(\frac{2}{q}\right)^\alpha$$

for $-q/2 < m \leqslant q/2$, then in a way similar to the proof of Lemma 4.10, we have

$$\sum_{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\, q)}{}' \frac{1}{\|\mathbf{m}\|^\alpha} - \sum_{\substack{(\mathbf{a},\mathbf{m}^{(0)})\equiv 0(\mathrm{mod}\, q)\\ -\frac{q}{2}<m_i^{(0)}\leqslant\frac{q}{2}}}{}' \frac{1}{\|\mathbf{m}^{(0)}\|^\alpha}$$

$$\leqslant \sum_{v=1}^{s} \sum_{\substack{(\mathbf{a},\mathbf{m}^{(0)})\equiv 0(\bmod q)\\ -\frac{q}{2}<m_i^{(0)}\leqslant\frac{q}{2}}} \left(\sum_{l_v=-\infty}^{\infty}{}' \frac{1}{\overline{(m_v^{(0)}+l_vq)}^{\alpha}}\right) \prod_{\substack{\mu=1\\ \mu\neq v}}^{s} \left(\sum_{l_\mu=-\infty}^{\infty} \frac{1}{\overline{(m_\mu^{(0)}+l_\mu q)}^{\alpha}}\right)$$
$$\leqslant s2^{\alpha}(2\zeta(\alpha)+1)^s q^{-\alpha}.$$

The lemma is proved.

**Lemma 8.3** *Suppose that $r$ is a positive integer. If the congruence*

$$(\mathbf{a},\mathbf{m}) = \sum_{i=1}^{s} a_i m_i \equiv 0(\bmod n) \tag{8.2}$$

*has no solution in the domain*

$$\|\mathbf{m}\| \leqslant M, \quad \mathbf{m}\neq \mathbf{0},$$

*then*

$$\sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0(\bmod n)\\ |m_i|\leqslant rn}}{}' \frac{1}{\|\mathbf{m}\|} \leqslant c(s)M^{-1}(\ln rn)^s.$$

*Proof.* Let $T_{l,M}(l\geqslant 1)$ be the number of solutions of (8.2) in the domain

$$\|\mathbf{m}\| < lM, \quad \mathbf{m}\neq \mathbf{0}.$$

Then $T_{l,M}\leqslant c(s)l(\ln 3lM)^{s-1}$ (Cf. Lemma 3.5). Hence

$$\sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0(\bmod n)\\ |m_i|\leqslant rn}}{}' \frac{1}{\|\mathbf{m}\|} \leqslant \sum_{l=1}^{(rn)^s} \frac{T_{l+1,M}-T_{l,M}}{lM}$$
$$= M^{-1}\sum_{l=1}^{(rn)^s} T_{l+1,M}\left(\frac{1}{l}-\frac{1}{l+1}\right) + \frac{T_{(rn)^s+1,M}}{((rn)^s+1)M}$$
$$\leqslant c(s)M^{-1}\sum_{l=1}^{(rn)^s}\frac{(\ln 3lM)^{s-1}}{l} + c(s)M^{-1}(\ln 3rn)^{s-1}$$
$$\leqslant c(s)M^{-1}(\ln rn)^s.$$

The lemma is proved.

**Theorem 8.4** $\mathbf{b}_1 = (1, b_1, \cdots, b_1^{s-1})$ *is a good lattice point* mod $p_1$.

*Proof.* By Lemma 8.1,

$$B_2(\{x\}) = \{x\}^2 - \{x\} + \frac{1}{6} = \frac{1}{2\pi^2}\sum_{m=-\infty}^{\infty}{}' \frac{e^{2\pi imx}}{m^2}$$

and so

$$3(1-2\{x\})^2 = \frac{6}{\pi^2}\sum_{m=-\infty}^{\infty}{}' \frac{e^{2\pi imx}}{m^2} + 1 = \sum_{m=-\infty}^{\infty}\frac{e^{2\pi imx}}{\psi(m)}, \tag{8.3}$$

where

$$\psi(m) = \begin{cases} \frac{\pi^2}{6} m^2, & \text{if } m \neq 0, \\ 0, & \text{if } m = 0. \end{cases}$$

Since

$$\left(1 - 2\left\{\frac{(p_1 - k)z^v}{p_1}\right\}\right)^2 = \left(1 - 2\left\{\frac{-kz^v}{p_1}\right\}\right)^2 = \left(1 - 2\left\{\frac{kz^v}{p_1}\right\}\right)^2$$

by (8.3), so

$$\begin{aligned} H_1(z) &= \frac{3^s}{p_1}\left(1 + \sum_{k=1}^{\frac{p_1-1}{2}} \prod_{v=0}^{s-1}\left(1 - 2\left\{\frac{kz^v}{p_1}\right\}\right)^2 + \sum_{k=1}^{\frac{p_1-1}{2}} \prod_{v=0}^{s-1}\left(1 - 2\left\{\frac{(p_1 - k)z^v}{p_1}\right\}\right)^2\right) \\ &= \frac{3^s}{p_1}\sum_{k=1}^{p_1} \prod_{v=0}^{s-1}\left(1 - 2\left\{\frac{kz^v}{p_1}\right\}\right)^2 \\ &= \frac{1}{p_1}\sum_{k=1}^{p_1} \sum \frac{e^{2\pi i(\mathbf{z},\mathbf{m})k/p_1}}{\prod_{v=1}^{s} \psi(m_v)} \\ &= 1 + \sum_{(\mathbf{z},\mathbf{m})\equiv 0(\bmod p_1)}{}' \frac{1}{\prod_{v=1}^{s}\psi(m_v)} < 1 + \sum_{(\mathbf{z},\mathbf{m})\equiv 0(\bmod p_1)}{}' \frac{1}{\|\mathbf{m}\|^2} \end{aligned}$$

and so

$$\begin{aligned} H_1(b_1) - 1 &\leqslant \min_{1 \leqslant z \leqslant p_1 - 1} \sum_{(\mathbf{z},\mathbf{m})\equiv 0(\bmod p_1)}{}' \frac{1}{\|\mathbf{m}\|^2} \\ &\leqslant c(s) p_1^{-2} (\ln p_1)^{2(s-1)} \end{aligned} \tag{8.4}$$

(Cf. Theorem 7.15). On the other hand

$$H_1(b_1) - 1 \geqslant \left(\frac{6}{\pi}\right)^s \sum_{(\mathbf{b}_1,\mathbf{m})\equiv 0(\bmod p_1)}{}' \frac{1}{\|\mathbf{m}\|^2}.$$

Therefore it follows that there exists $c(s)$ such that the congruence

$$(\mathbf{b}_1, \mathbf{m}) \equiv 0(\bmod p_1)$$

has no solution in the domain

$$\|\mathbf{m}\| < c(s) p_1 (\ln p_1)^{-(s-1)}, \quad \mathbf{m} \neq \mathbf{0}$$

and so the set

$$\left(\left\{\frac{k}{p_1}\right\}, \left\{\frac{kb_1}{p_1}\right\}, \ldots, \left\{\frac{kb_1^{s-1}}{p_1}\right\}\right), \quad 1 \leqslant k \leqslant p_1$$

has discrepancy

$$D(p_1) \leqslant c(s)p^{-1}(\ln p_1)^{2s-1}$$

by Theorem 3.2, i.e., $\mathbf{b}_1$ is a good lattice point mod $p_1$. The theorem is proved.

**Theorem 8.5** $(p_1 + p_2, p_1 b_2 + p_2 b_1, \cdots, p_1 b_2^{s-1} + p_2 b_1^{s-1})$ *is a good lattice point* mod $p_1 p_2$.

*Proof.* By (8.3), we have

$$H_2(z) = \frac{3^s}{p_1 p_2} \sum_{k=1}^{p_1 p_2} \prod_{v=0}^{s-1} \left(1 - 2\left\{\frac{(p_1 z^v + p_2 b_1^v)k}{p_1 p_2}\right\}\right)^2$$

and

$$H_2(z) - 1 = \frac{1}{p_1 p_2} \sum_{k=1}^{p_1 p_2} {\sum}' \frac{e^{2\pi i (p_1(\mathbf{z},\mathbf{m}) + p_2(\mathbf{b}_1,\mathbf{m}))k/p_1 p_2}}{\prod_{v=1}^{s} \psi(m_v)}$$

$$= \sum_{p_1(\mathbf{z},\mathbf{m})+p_2(\mathbf{b}_1,\mathbf{m}) \equiv 0 (\bmod p_1 p_2)}{}' \frac{1}{\prod_{v=1}^{s} \psi(m_v)}$$

$$= \sum_{\substack{(\mathbf{b}_1,\mathbf{m}) \equiv 0 (\bmod p_1) \\ (\mathbf{z},\mathbf{m}) \equiv 0 (\bmod p_2)}}{}' \frac{1}{\prod_{v=1}^{s} \psi(m_v)}.$$

Divide the last sum into two parts $\Sigma_1$ and $\Sigma_2$, where $\Sigma_1$ contains those $\mathbf{m}$ such that $p_2 | m_v$ for $1 \leqslant v \leqslant s$ and $\Sigma_2$ contains the remaining terms. By (8.4) and Lemma 8.2, we have

$$\sum\nolimits_1 = \sum_{p_2(\mathbf{b}_1,\mathbf{m}) \equiv 0 (\bmod p_1)}{}' \frac{1}{\prod_{v=1}^{s} \psi(p_2 m_v)}$$

$$\leqslant p_2^{-2} \sum_{(\mathbf{b}_1,\mathbf{m}) \equiv 0 (\bmod p_1)}{}' \frac{1}{\prod_{v=1}^{s} \psi(m_v)}$$

$$= p_2^{-2}(H_1(b_1) - 1) \leqslant c(s)(p_1 p_2)^{-2}(\ln p_1 p_2)^{2(s-1)}$$

and

$$\sum\nolimits_2 \leqslant \sum_{\substack{(\mathbf{b}_1,\mathbf{m}) \equiv 0 (\bmod p_1) \\ (\mathbf{z},\mathbf{m}) \equiv 0 (\bmod p_2)}}{}_2 \frac{1}{\|\boldsymbol{m}\|^2}$$

$$\leqslant \sum_{-\frac{p_1 p_2}{2} < m_i \leqslant \frac{p_1 p_2}{2}}{}_2 \sum_{\substack{(\mathbf{b}_1,\mathbf{m}) \equiv 0 (\bmod p_1) \\ (\mathbf{z},\mathbf{m}) \equiv 0 (\bmod p_2)}} \frac{1}{\|\mathbf{m}\|^2} + c(s)(p_1 p_2)^{-2}$$

$$\leqslant \left( \sum\nolimits_{2}_{-\frac{p_1p_2}{2}<m_i\leqslant\frac{p_1p_2}{2}} \sum_{\substack{(\mathbf{b_1},\mathbf{m})\equiv 0(\bmod p_1)\\ (\mathbf{z},\mathbf{m})\equiv 0(\bmod p_2)}} \frac{1}{\|\mathbf{m}\|} \right)^2 + c(s)(p_1p_2)^{-2}.$$

Hence it follows by Theorem 8.4 and Lemmas 4.5 and 8.3 that

$$\begin{aligned}
\min_{1\leqslant z\leqslant\frac{p_2-1}{2}} \sum\nolimits_2 &\leqslant \left( \min_{1\leqslant z\leqslant\frac{p_2-1}{2}} \sum\nolimits_{2}_{\frac{-p_1p_2}{2}<m_i\leqslant\frac{p_1p_2}{2}} \sum_{\substack{(\mathbf{b_1},\mathbf{m})\equiv 0(\bmod p_1)\\ (\mathbf{z},\mathbf{m})\equiv 0(\bmod p_2)}} \frac{1}{\|\mathbf{m}\|} \right)^2 + c(s)(p_1p_2)^{-2} \\
&\leqslant \left( \frac{2}{p_2-1} \sum\nolimits_{2}_{\substack{\frac{-p_1p_2}{2}<m_i\leqslant\frac{p_1p_2}{2}\\ (\mathbf{b_1},\mathbf{m})\equiv 0(\bmod p_1)}} \frac{1}{\|\mathbf{m}\|} \sum_{\substack{1\leqslant z\leqslant\frac{p_2-1}{2}\\ (\mathbf{z},\mathbf{m})\equiv 0(\bmod p_2)}} 1 \right)^2 + c(s)(p_1p_2)^{-2} \\
&\leqslant \left( \frac{2(s-1)}{p_2-1} \sideset{}{'}\sum_{\substack{\frac{-p_1p_2}{2}<m_i\leqslant\frac{p_1p_2}{2}\\ (\mathbf{b_1},\mathbf{m})\equiv 0(\bmod p_1)}} \frac{1}{\|\mathbf{m}\|} \right)^2 + c(s)(p_1p_2)^{-2} \\
&\leqslant c(s)(p_1p_2)^{-2}(\ln p_1p_2)^{4s-2}.
\end{aligned}$$

Consequently.

$$H_2(b_2)-1 \leqslant \Sigma_1 + \min_{1\leqslant z\leqslant\frac{p_2-1}{2}} \Sigma_2 \leqslant c(s)(p_1p_2)^{-2}(\ln p_1p_2)^{4s-2}.$$

On the other hand,

$$H_2(b_2)-1 \geqslant \left(\frac{b}{\pi^2}\right)^s \sideset{}{'}\sum_{p_1(\mathbf{b_2},\mathbf{m})+p_2(\mathbf{b_1},\mathbf{m})\equiv 0(\bmod p_1p_2)} \frac{1}{\prod_{v=1}^{s}\psi(m_v)}.$$

It follows that there exists $c(s)$ such that the congruence

$$p_1(\mathbf{b_2},\mathbf{m})+p_2(\mathbf{b_1},\mathbf{m}) \equiv 0(\bmod p_1p_2)$$

has no solution in the domain

$$\|\mathbf{m}\| \leqslant c(s)p_1p_2(\ln p_1p_2)^{-(2s-1)}, \quad \mathbf{m}\neq\mathbf{0}.$$

Hence the theorem follows by Theorem 3.2.

Similarly, we may prove

**Theorem 8.6** *Let* $e_l = q_1b_1^{l-1}+\cdots+q_tb_t^{l-1}(1\leqslant l\leqslant s)$. *Then* $(e_1,\cdots,e_s)$ *is a good lattice point* mod $q$.

Let $x$ denote the solution of the congruence

$$(q_1 + \cdots + q_t)x \equiv 1(\text{mod} q), \quad 1 \leqslant x \leqslant q.$$

Let

$$h_{v+1} \equiv (q_1 b_1^v + \cdots + q_t b_t^v)x(\text{mod} q),$$
$$1 \leqslant h_{v+1} \leqslant q, \quad 1 \leqslant v \leqslant s-1.$$

Then the good lattice point given by theorem 8.6 may be written as

$$(1, h_1, \cdots, h_s).$$

*Remark*. P. Keast [1, 2] has pointed out that the error term of the quadrature formula corresponding to a good lattice point given by Theorem 8.6 is comparatively small, if we take $p_1, p_2, \cdots$ and $p_t$ almost equal, when $t$ is confined by $t \leqslant 4$. It is also suggested by Theorem 8.5 that if we have a good lattice point mod $n$ and a prime $p$ which is not a divisor of $n$, then we may obtain a good lattice point mod $pn$.

## 8.7 Several remarks

1. We have seen from §2 − §4 that the precision of the error terms of the quadrature formulas given by $\mathscr{R}_s$ set and *glp* set are better than other quadrature formulas and another advantage is that the so obtained quadrature formula has a very simple form, i.e., the arithmatic mean of the values of the integrand at the given set of points is used to approximate the definite integral.

2. It follows from Theorem 8.4 that

$$c(s)n^2$$

elementary operations are required to obtain a good lattice point mod $n$, where $n = p_1$. Let $p_2 \doteq \sqrt{p_1}$ and $n = p_1 p_2$. Then by Theorem 8.5, the number of elementary operations for obtaining a good lattice point mod $n$ is

$$c(s)(p_1^2 + p_1 p_2^2) = c(s)(p_1 p_2)^{4/3} = c(s)n^{4/3}.$$

In general, we may take $p_{v+1} \doteq \sqrt{p_v}(1 \leqslant v \leqslant t-1)$ and $n = p_1 \cdots p_t$. Then the number of elementary operations for obtaining a good lattice point mod $n$ are

$$c(s)(p_1^2 + p_1 p_2^2 + \cdots + p_1 \cdots p_{t-1} p_t^2) = c(s)n^{\frac{2}{2-2^{-(t-1)}}}. \tag{8.5}$$

If it is required that the error term of the corresponding quadrature formula should be comparatively small, then it is better to take the $p_i's$ almost equal according to the opinion of $P$. Keast [1, 2] and so (8.5) should be replaced by

$$c(s)n^{1+t^{-1}}.$$

However it requires only

$$O(\ln n)$$

elementary operations for obtaining the $\eta$ set or $\mathscr{R}_s$ set, where the constant implied by the symbol "$O$" depends on $\eta$ or $\mathscr{R}_s$ only. Of course, $O(n)$ elementary operations are still needed for obtaining the corresponding $W_2(n, \mathbf{h})$ of $\eta$ set or $\mathscr{R}_s$ set.

As for the comparison of *glp* set and $\mathscr{R}_s$ set, we quote the opinion expressed by S. Haber [2] as follows.

"The second method takes $AN^2s$ seconds (to get $a^*$ $(N, s')$ and $B^*(N, s')$ for all $s' \leqslant s$), where now $A$ was 0.00001 for the calculations reported here (on a UNIVAC 1108). This is very much better, and the method can be carried out at reasonable expense for $N$ up to 10,000 or so and $s$ up to 10 or 20. However, there is reason to suspect that practical formulas for $s$ as high as 10 will reguire $N's$ of order 100,000 or more, and again the calculation becomes excessively long.

The third method requires only $As^3$ seconds, the only calculation of any significant length that is necessary is the solution of the linear system (12). $A$ is apt to be about $10^{-3}$ or lower. The length of calculation is thus no obstacle for $s$ up to 100, at least, and $N$ abitrary large. While it is not known that this method actually produces g.l.p. sequences, the numerical evidence indicates that it does. However, the evidence also indicates that the quadrature formulas produced have error bounds much higher than do those produced by the first two methods."

Here the second method refers to the method given by Theorem 8.4 of which $a^*(N, s')$ and $B^*(N, s')$ are $(1, b_1, \cdots, b_1^{s-1})$ mod $N$ and $W_2(N, \mathbf{b}_1)$ respectively. The third method means the method stated in §8.4 and the linear system (12) is (8.1).

The calculation of Y. S. Moon [1] given by an IBM 370/165 leads to a similar conclusion.

3. As for the comparison of the $W_2(n, \mathbf{h})$ given by *glp* set and $\mathscr{R}_s$ set, we may see $M$. Maisonneuve [1] and S. Haber [2] for the cases $s = 3, 5, 6$. Now we give some data for the cases $s = 7, 8, 9$ as follows.

**Table 1** $(s = 7)$

| $\mathscr{R}_s$ set | | *glp* set | |
|---|---|---|---|
| $n$ | $W_2(n, \mathbf{h})$ | $n$ | $W_2(n, \mathbf{h})$ |
| 11,215 | 1.9416 | 15,019 | 1.2 |
| 84,523 | $2.0407 \times 10^{-1}$ | 71,053 | $2.1 \times 10^{-1}$ |

**Table 2** $(s = 8)$

| $\mathscr{R}_s$ set | | *glp* set | |
|---|---|---|---|
| $n$ | $W_2(n, \mathbf{h})$ | $n$ | $W_2(n, \mathbf{h})$ |
| 28,832 | $3.4501 \times 10^{-1}$ | 24,041 | 3.9 |
| 84,523 | $9.8761 \times 10^{-1}$ | 100,063 | $7.6 \times 10^{-1}$ |

**Table 3** $(s = 9)$

| $\mathscr{R}_s$ set | | *glp* set | |
|---|---|---|---|
| $n$ | $W_2(n, \mathbf{h})$ | $n$ | $W_2(n, \mathbf{h})$ |
| 42,570 | $1.0496 \times 10$ | 46,213 | 9.5 |
| 172,155 | 2.3708 | 159,053 | 2.5 |

4. The $W_2(n, \mathbf{h})$ given by $\mathscr{R}_s$ set is much better than that given by $\eta$ set, especially when $s$ is comparatively large. For example, we have

$$W_2(957, 833, \mathbf{h}) \leqslant 4.1494 \times 10^{-1}$$

from the $\mathscr{R}_s$ set and

$$W_2(1, 035, 269, \mathbf{h}) \leqslant 4.6013 \times 10^{-1}$$

from $\eta$ set for $s = 9$, and

$$W_2(7, 494, 007, \mathbf{h}) \leqslant 6.3956 \times 10^{-1}$$

from $\mathscr{R}_s$ set and

$$W_2(8, 359, 937, \mathbf{h}) \leqslant 1.0401$$

from $\eta$ set for $s = 11$.

5. So far as we know, the cyclotomic field often gives the most precise results in applications to the problems of mumerical analysis among the real algebraic number fields of the same degree and its other advantage is the convenience for computation.

## 8.8 Tables

The tables given in the Appendix contain the $n, \mathbf{h}(n),\ \rho(n, \mathbf{h}),\ W_2(n, \mathbf{h})$ and $W_4(n, \mathbf{h})$ of various dimensions. It is not only necessary for the approximate evaluation of multiple integral but also important for theoretical work in numerical analysis. The most useful table of A. I. Saltykov [1] was made according to Korobov's method (Cf. §8.6). It is confined only to $3 \leqslant s \leqslant 10$ and $100 \leqslant n \leqslant 155,093$, since it requires long calculations for obtaining good lattice points. He also gave the upper estimate of $\sup\limits_{f \in E_s^2(C)} |S(n, \mathbf{h}, f)|$ which is rougher than the function of $W_2(n, \mathbf{h})$. Saltykov's

table is contained in many monographs (Cf. N. M. Korobov [7], Hua Loo Kang and Wang Yuan [3] and A. H. Stroud [1]). By the use of a CDC 6400, M. Maisonneuve [1] published the calculation of $W_2(n, \mathbf{h})$ of the pairs $(n, \mathbf{h})$ contained in Saltykov's table. H. Conroy [1] gave several tables of good lattice points of $s \leqslant 12$ and P. Keast [1, 2] gave some data obtained by the use of Theorem 8.6. By a more complicated but precise method, M. Maisonueuve [1], G. Kedem and S. K. Zaremba [1] gave some data for $(n, \mathbf{h})$, $W_2(n, \mathbf{h})$ and $W_4(n, \mathbf{h})$ for $s = 3.4$ and $n \leqslant 6.606$. Recently, it was pointed out by R. Cranley and T. N. L. Patterson [1]:

"The number of rules available, particular with large number of points, is also very limited, the most extensive table been Saltykov and Conroy, Further rule would have to be computed although this task would be easier by the work of Hua and Wang."

The first $\mathscr{R}_s$ sets were given by Hua Loo Keng and Wang Yuan [4, 5] of which one is $s = 11$ and $n = 698,047$. Later, S. Haber [2] and Y. S. Moon [1] gave some tables of $\mathscr{R}_s$ sets of $s \leqslant 14$ and $n \leqslant 10^6$ by the use of computers UNIVAC 1108 and IBM 370/165 respectively. Recently Wang Yuan, Xu Guang Shan and Zhang Rong Xiao [1] published a more extended table of $\mathscr{R}_s$ set of $s \leqslant 18$ by the use of Djs-013.

Table 1 and tables 10–12 in the Appendix are given by $\mathscr{R}_s$. In tables 2–9, the data marked with star are given by $\eta$ or $\mathscr{R}_s$ and the others are good lattice points given by Korobov's method.

## 8.9 Some examples

Suppose that $D$ is a bounded domain with piecewise smooth boundary. If $\mathscr{D} = G_s$, then we may use a simple sum constructed by the values of integrand at the given uniformly distributed set of points to approximate the definite integral

$$\int_{\mathscr{D}} f(\mathbf{x}) d\mathbf{x}$$

(Cf. §5.4). If $D$ is not $G_s$, then without loss of generality, we may suppose that $\mathscr{D} \subset G_s$. Let

$$F(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \text{if } \mathbf{x} \in \mathscr{D} \\ 0 & \text{if } \mathbf{x} \overline{\in} \mathscr{D}. \end{cases}$$

Then

$$\int_{\mathscr{D}} f(\mathbf{x}) d\mathbf{x} = \int_{G_s} f(\mathbf{x}) d\mathbf{x}.$$

Now we give some examples as follows.

1. Denote

$$I = 4\int_0^1\int_0^1 x_1x_2dx_1dx_2 = 1.$$

Let $I_1$ be the approximate value of $I$ given by the quadrature formula constructed by $\eta$ set of dimension 2 (Cf. example 3 of §8.3). Further let $I_2, I_3$ and $I_4$ denote the approximate values of $I$ given also by the quadrature formula of which the integrand are obtained by changing variables

$$x_i = y_i^2(1-y_i),$$
$$x_i = y_i^3(10-15y_i+6y_i^2)$$

and

$$x_i = y_i^4(35-84y_i+70y_i^2-20y_i^3), \quad i=1,2$$

respectively (§6.5). Then we have

**Table 1**

| $n$ | $I_1$ | $I_2$ | $I_3$ | $I_4$ |
|---|---|---|---|---|
| 13 | $9.2308\times10^{-1}$ | 1.0285 | 1.0022 | $9.9781\times10^{-1}$ |
| 21 | $9.4029\times10^{-1}$ | 1.0131 | 1.0023 | 1.0004 |
| 34 | $9.7059\times10^{-1}$ | 1.0060 | 1.0006 | 1.00002 |
| 55 | $9.7709\times10^{-1}$ | 1.0027 | 1.0001 | 1.000014 |

2. Denote

$$\begin{aligned}
J_1 &= \int_0^1\int_0^1 \frac{x_1^2}{1+x_2^2}dx_1dx_2 = 2.6179\cdots\times10^{-1},\\
J_2 &= \int_0^1\int_0^1\int_0^1(1+3x_1x_2x_3+x_1^2x_2^2x_3^2)e^{x_1x_2x_3}dx_1dx_2dx_3\\
&= 1.7182\cdots,\\
J_3 &= \int_{-1}^1\int_{-1}^1\int_{-1}^1 \exp(\sin x_1\sin x_2\sin x_3)dx_1dx_2dx_3\\
&= 8.0817\cdots,\\
J_4 &= \int_0^2\int_{x_1^2}^{2x_1} x_1x_2^2dx_1dx_2 = 6.4
\end{aligned}$$

and

$$J_5 = \int_0^1\int_{x_1}^{2x_1}\int_{x_1x_2}^{x_1^2x_2} x_1^3x_2^3x_3dx_1dx_2dx_3 = -4.3356\cdots\times10^{-2}.$$

Let $J_i'(1\leqslant i\leqslant 5)$ denote the approximate values of $J_i(1\leqslant i\leqslant5)$ given by the quadrature formula constructed by $\mathscr{R}_3$ set (Cf. example 2 of §8.4) respectively. Then we have

**Table 2**

| $n$ | $J'_1$ | $J'_2$ | $J'_3$ | $J'_4$ | $J'_5$ |
|---|---|---|---|---|---|
| 20 | $3.5573 \times 10^{-1}$ | 1.4992 | 8.9476 | 8.0148 | $-6.7248 \times 10^{-2}$ |
| 83 | $2.9544 \times 10^{-1}$ | 1.6932 | 7.8366 | 6.9748 | $-5.2348 \times 10^{-2}$ |
| 418 | $2.6103 \times 10^{-1}$ | 1.7202 | 7.9947 | 6.3749 | $-4.2538 \times 10^{-2}$ |
| 1,692 | $2.6180 \times 10^{-1}$ | 1.7183 | 8.0844 | 6.3999 | $-4.3290 \times 10^{-2}$ |
| 3,802 | $2.6180 \times 10^{-1}$ | 1.7183 | 8.0817 | 6.4000 | $-4.3357 \times 10^{-2}$ |

3. Let

$$K = \int_{-1}^{1}\int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}}\int_{-\sqrt{1-x_1^2-x_2^2}}^{\sqrt{1-x_1^2-x_2^2}} \frac{dx_1dx_2dx_3}{x_1^2+x_2^2+(x_3+0.5)^2}$$
$$= 1.1460\cdots \times 10.$$

Now we give a comparison of the precisions of the approximate values of $K$ given by the quadrature formula constructed by the $\mathscr{R}_3$ set (Cf. example 2 of §8.4) and the Cartesion product formula of Gaussian quadrature formula as follows.

**Table 3**

| Methods | $n$ | Approximate values of $K$ | Times in Djs-6 |
|---|---|---|---|
| Gauss Method | $2^3$ | $1.2409 \times 10$ | 2.6 sec. |
| | $4^3$ | $1.1324 \times 10$ | 15 sec. |
| | $8^3$ | $1.1391 \times 10$ | 104 sec. |
| | $16^3$ | $1.1425 \times 10$ | 789 sec. |
| Number Theoretic method | 20 | $1.3063 \times 10$ | 0.5 sec. |
| | 83 | $1.0928 \times 10$ | 1 sec. |
| | 418 | $1.1494 \times 10$ | 9 sec. |
| | 1,692 | $1.1460 \times 10$ | 41 sec. |

*Remarks.* 1. The precision of numerical integration is often higher, if the integrand is replaced by a periodic function (Cf. §6.4–§6.5). But the integrand becomes complicated, if it has changed its variables. Hence careful analysis of the integrand is needed in practical use (Cf. D. Maisoneuve [1]).

2. If s is large, the integral over $G_s$ may be divided into several parts and then the quadrature formula of lower dimensions can be applied to each part respectively. For example, suppose that $s = s' + s''$ and the good lattice points of $s'$ and $s''$ dimensions are $\mathbf{h}'$ (mod $n'$) and $\mathbf{h}'$ (mod $n''$) respectively. Then

$$\frac{1}{n'n''}\sum_{k=1}^{n'}\sum_{l=1}^{n''} f\left(\frac{k\mathbf{h}'}{n'}, \frac{l\mathbf{h}''}{n''}\right)$$
$$= \frac{1}{n'n''}\sum_{k=1}^{n'}\sum_{l=1}^{n''}\sum C(\mathbf{m})e^{2\pi i(\frac{(\mathbf{h}',\mathbf{m}')k}{n'}+\frac{(\mathbf{h}'',\mathbf{m}'')l}{n''})}$$

$$= \sum_{\substack{(\mathbf{h}',\mathbf{m}')\equiv 0 (\mathrm{mod}\ n') \\ (\mathbf{h}'',\mathbf{m}'')\equiv 0 (\mathrm{mod}\ n'')}} C(\mathbf{m}),$$

where $\mathbf{m}=(\mathbf{m}',\mathbf{m}'')$ in which $\mathbf{m}'$ and $\mathbf{m}''$ denote the integral vectors of s′ and s″ dimensions respectively. Hence

$$\sup_{f\in E_s^\alpha(c)} \left| \int_{G_s} f(\mathbf{x})d\mathbf{x} - \frac{1}{n'n''}\sum_{k=1}^{n'}\sum_{l=1}^{n''} f\left(\frac{k\mathbf{h}'}{n'}, \frac{l\mathbf{h}''}{n''}\right) \right|$$
$$\leqslant C \sum_{\substack{(\mathbf{h}'\mathbf{m}')\equiv 0 (\mathrm{mod}\ n') \\ (\mathbf{h}'',\mathbf{m}'')\equiv 0 (\mathrm{mod}\ n'')}}{}' \frac{1}{(\|\mathbf{m}'\|\|\mathbf{m}''\|)^\alpha}.$$

It follows by the well-known method that the right hand side is dominated by

$$O((\min(n',n''))^{-\alpha}(\ln n'n'')^{\alpha(s-1)})$$

(Cf. §7.6). In general, the error is comparatively large. But some advan-tages may be obtained for some particular values of $n = n'n''$ (Cf. S. K. Zaramba [3]).

**Notes**

Theorems 8.2 and 8.3: Cf. Zhang Roug Xiao [1], S. Haber [2] and M. Maisoneuve [1], §2–§3: Cf. Hua Loo Keng and Wang Yuan [6,7].

§4–§5: Cf. Hua Loo Keng and Wang Yuan [1,4,5,6,7]. S. Haber, [2], Y. S. Moon [1] and Wang Yuan, Xu Guang Shan and Zhang Rong Xiao [1].

Theorems 6.4 and 6.5 were proved by N. M. Korobov [5,7] and Theorem 6.6 was obtained by Wang Yuan, Zhu Yao Cheng and Jian Yun Cui [1] and P. Keast [1,2].

Concerning the numerical integration over a domain $\mathscr{D} \neq G_s$, we may refer also V. M. Colodov [2].

The examples of §2 and §9 were given by Xu Fong by the use of Djs–6.

# Chapter 9

# Interpolation

## 9.1 Introduction

Let $1 < n_1 < n_2 < \cdots$ be a sequence of integers and let

$$P_{n_l}(k) = (x_1^{(n_l)}(k), \cdots, x_s^{(n_l)}(k)), \quad 1 \leqslant k \leqslant n_l, l = 1, 2, \cdots$$

be a sequence of uniformly distributed sets in $G_s$. For any given function f($\mathbf{x}$) on $G_s$, let

$$P_f(\mathbf{x}) = \sum_{k=1}^{n_l} f(P_{n_l}(k))\psi_{n_l,k}(\mathbf{x}), \tag{9.1}$$

where $\psi_{n_l,k}(\mathbf{x})(1 \leqslant k \leqslant n_l)$ are given functions.

If $P_f(\mathbf{x})$ converges to $f(\mathbf{x})$ according to a certain measure as $n_l \to \infty$, then $P_f(\mathbf{x})$ is called the approximate polynomial of $f(\mathbf{x})$. For simplicity, we use $P(\mathbf{x})$ or $P$ instead of $P_f(\mathbf{x})$.

The measures that are often used are as follows.

1. Absolute error

$$\sup_{\mathbf{x} \in G_s} |P(\mathbf{x}) - f(\mathbf{x})|.$$

2. Mean square error

$$\|P - f\|_2 = \left( \int_{G_s} |P - f|^2 d\mathbf{x} \right)^{1/2}.$$

Suppose that $f$ has an absolutely convergent Fourier expansion

$$f(\mathbf{x}) = \Sigma C(\mathbf{m}) e^{2\pi i(\mathbf{m},\mathbf{x})},$$

where

$$C(\mathbf{m}) = \int_{G_s} f(\mathbf{x}) e^{-2\pi i(\mathbf{m},\mathbf{x})} d\mathbf{x}.$$

Then by the method of numerical integration, we may use

$$\sum_{k=1}^{n_l} \alpha_{n_l,k} f(P_{n_l}(k)) e^{-2\pi i(\mathbf{m}, P_{n_l}(k))}$$

to approximate the Fourier coefficient $C(\mathbf{m})$ and so we may expect to use

$$P(\mathbf{x}) = \sum_{\|\mathbf{m}\|<N(n_l)} \sum_{k=1}^{n_l} \alpha_{n_l,k} f(P_{n_l}(k)) e^{2\pi i(\mathbf{m}, \mathbf{x}-P_{n_l}(k))}$$

to approximate $f(\mathbf{x})$. This is the simplest method to construct the approximate polynomial. We may also use other methods to construct the approximate polynomials. The results of this chapter are generalization and application of the multiple quadrature stated in previous chapters.

## 9.2 The set of equi-distribution and interpolation

Let $m$ be an integer $\geqslant 2$, $n = m^s$, $\mathbf{I} = (l_1, \cdots, l_s)$ and

$$P(\mathbf{x}) = \frac{1}{n} \sum_{\substack{0 \leqslant l_i < m \\ 1 \leqslant i \leqslant s}} f\left(\frac{\mathbf{I}}{m}\right) \sum_{\|\mathbf{k}\|<N} e^{2\pi i(\mathbf{k}, \mathbf{x}-\frac{\mathbf{I}}{m})}.$$

**Theorem 9.1** *Suppose that* $\alpha > 1$ *and* $N = \left[\frac{m}{2}\right]$. *Then*

$$\sup_{f \in E_s^\alpha(C)} \|P - f\|_2 \leqslant C_c(\alpha, s) n^{-\frac{2\alpha-1}{2s}} (\ln n)^{\frac{s-1}{2}}.$$

*Proof.* Suppose that $f \in E_s^\alpha(C)$ and

$$f(\mathbf{x}) = \Sigma C(\mathbf{m}) e^{2\pi i(\mathbf{m}, \mathbf{x})}.$$

Then

$$\begin{aligned} & C(\mathbf{m}) - \frac{1}{n} \sum_{\substack{0 \leqslant l_i < m \\ 1 \leqslant i \leqslant s}} f\left(\frac{1}{m}\right) e^{-2\pi i(\mathbf{m}, \frac{\mathbf{I}}{m})} \\ &= C(\mathbf{m}) - \frac{1}{n} \sum_{\substack{0 \leqslant l_i < m \\ 1 \leqslant i \leqslant s}} \sum C(\mathbf{k}) e^{2\pi i(\mathbf{k}-\mathbf{m}, \frac{\mathbf{I}}{m})} \\ &= C(\mathbf{m}) - \sum_{\substack{m|(k_i - m_i) \\ 1 \leqslant i \leqslant s}} C(\mathbf{k}) \\ &= -\sum_{\substack{m|r_i \\ 1 \leqslant i \leqslant s}}' C(\mathbf{r}+\mathbf{m}). \end{aligned}$$

Since for integer $t$,

$$\int_0^1 e^{2\pi itx}dx = \begin{cases} 1, & \text{if } t = 0 \\ 0, & \text{if } t \neq 0, \end{cases}$$

therefore

$$\begin{aligned} \|P - f\|_2^2 &= \int_{G_s} |P - f|^2 d\mathbf{x} \\ &= \int_{G_s} (P - f)\overline{(P - f)} d\mathbf{x} \\ &= \sum_{\|\mathbf{m}\|<N} \left| \sum_{\substack{m|r_i \\ 1\leqslant i\leqslant s}}{}' C(\mathbf{r} + \mathbf{m}) \right|^2 + \sum_{\|\mathbf{m}\|\geqslant N} |C(\mathbf{m})|^2. \end{aligned} \tag{9.2}$$

Hence

$$\sup_{f\in E_s^\alpha(C)} \|P - f\|_2 \leqslant C^2(\Sigma_1 + \Sigma_2),$$

where

$$\Sigma_1 = \sum_{\|\mathbf{m}\|<N} \left( \sum_{\substack{m|r_i \\ 1\leqslant i\leqslant s}}{}' \frac{1}{\|\mathbf{r} + \mathbf{m}\|^\alpha} \right)^2$$

and

$$\Sigma_2 = \sum_{\|\mathbf{m}\|\geqslant N} \frac{1}{\|\mathbf{m}\|^{2\alpha}}.$$

By Lemma 7.7, we have

$$\begin{aligned} \Sigma_1 &\leqslant s \sum_{\|\mathbf{m}\|<N} \left( 2\sum_{r=1}^{\infty} \frac{1}{m^\alpha \left(r - \frac{1}{2}\right)^\alpha} \right)^2 \left( 1 + 2\sum_{r=1}^{\infty} \frac{1}{m^\alpha \left(r - \frac{1}{2}\right)^\alpha} \right)^{2(s-1)} \\ &\leqslant c(\alpha, s)m^{-2\alpha} \sum_{\|\mathbf{m}\|<N} 1 \leqslant c(\alpha, s)m^{-2\alpha+1}(\ln n)^{s-1} \end{aligned}$$

and

$$\Sigma_2 \leqslant c(\alpha, s)m^{-2\alpha+1}(\ln n)^{s-1}.$$

The theorem follows.

**Theorem 9.2** *Suppose that* $\alpha > 1$. *Then*

$$\sup_{f\in E_s^\alpha(C)} \|P - f\|_2 \geqslant \begin{cases} C, & \text{if } N > m, \\ \dfrac{C}{\sqrt{2\alpha - 1}} n^{-\frac{2\alpha-1}{2s}}, & \text{if } N \leqslant m. \end{cases}$$

*Proof.* Take

$$f(\mathbf{x}) = \sum_{k=-\infty}^{\infty} \frac{e^{2\pi i k x_1}}{\bar{k}^{\alpha}}.$$

Then by (9.2),

$$\|P - f\|_2^2 = C^2 \sum_{\bar{k}<N} \left( \sum_{m|r}{}' \frac{1}{(\overline{r+k})^{\alpha}} \right)^2 + C_2 \sum_{\bar{k}\geqslant N} \frac{1}{\bar{k}^{2\alpha}}.$$

Hence

$$\|P - f\|_2^2 \geqslant C^2 \sum_{\bar{k}<N} \frac{1}{(\overline{m+k})^{2\alpha}} \geqslant C^2$$

for $N > m$ and

$$\|P - f\|_2^2 \geqslant C^2 \sum_{k\geqslant N} \frac{1}{k^{2\alpha}} \geqslant C^2 \int_N^{\infty} \frac{dt}{t^{2\alpha}}$$
$$= \frac{C^2}{2\alpha - 1} N^{-2\alpha+1} \geqslant \frac{C^2}{2\alpha - 1} m^{-2\alpha+1}$$

for $N \leqslant m$. The theorem follows.

It follows from Theorem 9.2 that the principal order $n^{-\frac{2\alpha-1}{2s}}$ of the error term in Theorem 9.1 can not be improved further. In the following, we shall study the lower estimate of the error term between the function and any of its approximate polynomial constructed by the set of equi-distribution.

**Lemma 9.1** *Suppose that*

$$\psi(x) = \begin{cases} \left( \dfrac{\sin 2\pi m x}{2m} \right)^{\alpha-1}, & \text{if } 0 \leqslant x \leqslant \dfrac{1}{2m}, \\ 0, & \text{if } \dfrac{1}{2m} \leqslant x \leqslant 1 \end{cases}$$

*and* $\psi(x+1) = \psi(x)$, *where* $\alpha$ *is an integer* $> 1$. Then

$$f(\mathbf{x}) = \psi(x_1) \cdots \psi(x_s) \in E_s^{\alpha}(c(\alpha)^s).$$

*Proof.* Since

$$\psi(x) = \left( \frac{e^{2\pi i m x} - e^{-2\pi i m x}}{4im} \right)^{\alpha-1}$$
$$= \frac{1}{(4im)^{\alpha-1}} \sum_{\beta=0}^{\alpha-1} C_{\beta}^{\alpha-1} e^{2\pi i m(2\beta-\alpha+1)x},$$

therefore

$$\sup_{x\in G_1} |\varphi^{(\alpha-1)}(x)| \leqslant \frac{|2\pi i m(\alpha-1)|^{\alpha-1}2^{\alpha-1}}{|4im|^{\alpha-1}} = \pi^{\alpha-1}(\alpha-1)^{\alpha-1}$$

and

$$\sup_{x\in G_1} |\psi^{(\alpha)}(x)| \leqslant 2m\pi^{\alpha}(\alpha-1)^{\alpha}.$$

Since

$$\psi^{(v)}(0) = \psi^{(v)}\left(\frac{1}{2m}\right) = 0, \quad v = 0, 1, \cdots, \alpha-2,$$

so

$$\begin{aligned} C(k) &= \int_0^1 \psi(x)e^{-2\pi i k x}dx \\ &= \frac{1}{(2\pi i k)^{\alpha-1}} \int_0^{\frac{1}{2m}} \psi^{(\alpha-1)}(x)e^{-2\pi i k x}dx \\ &= \frac{-1}{(2\pi i k)^{\alpha}}\psi^{(\alpha-1)}(x)e^{-2\pi i k x}\Big|_0^{\frac{1}{2m}} + \frac{1}{(2\pi i k)^{\alpha}}\int_0^{\frac{1}{2m}} \psi^{(\alpha)}(x)e^{-2\pi i k x}dx \end{aligned}$$

and

$$|C(k)| \leqslant c(\alpha)\bar{k}^{-\alpha}.$$

The lemma is proved.

**Theorem 9.3** *Suppose that* $\alpha$ *is an integer* $> 1$ *and* $P(\mathbf{x})$ *is an approximate polynomial of* $f$ *of the type* (9.1) *defined by the set of equi-distribution* $1/m$. *Then*

$$\sup_{f\in E_s^{\alpha}(C)} \sup_{\mathbf{x}\in G_s} |f-P| \geqslant Cc(\alpha,s)n^{-\frac{\alpha-1}{s}}.$$

*Proof.* Take

$$f(\mathbf{x}) = Cc(\alpha,s)\psi\left(x_1 + \frac{1}{2m}\right),$$

where $\psi(x)$ is defined in Lemma 9.1. Then we may choose suitable $c(\alpha,s)$ such that $f \in E_s^{\alpha}(C)$. Since

$$f\left(\frac{\mathbf{l}}{m}\right) = Cc(\alpha,s)\psi\left(\frac{l_1}{m} + \frac{1}{2m}\right) = 0, \quad 0 \leqslant l_i \leqslant m, 1 \leqslant i \leqslant s,$$

so $P\left(\frac{\mathbf{l}}{m}\right) = 0$. On the other hand

$$f\left(1-\frac{1}{4m}, 0, \cdots, 0\right) = Cc(\alpha,s)\psi\left(\frac{1}{4m}\right)$$

$$= Cc(\alpha, s)m^{-\alpha+1} = Cc(\alpha, s)n^{-\frac{\alpha-1}{s}}.$$

The theorem is proved.

It follows from Theorems 9.2 and 9.3 that the error terms between function and its approximate polynomials constructed by using the set of equi-distribution are comparatively large. In the following, we shall use the $\mathscr{R}_s$ set and *glp* set to construct the approximate polynomial $P$ of the function $f$ such that the principal order of the error term between $P$ and $f$ is independent of $s$.

## 9.3 Several lemmas

**Lemma 9.2** *Let* $\mathrm{I} = (l_1, \cdots, l_s)$ *be an integral vector satisfying* $\|\mathrm{I}\| \geqslant 3^s$ *and let* $N$ *be a number satisfying* $1 \leqslant N \leqslant \|\mathrm{I}\|/3^s$. *Then*

$$\sum_{\|\mathbf{m}\| \leqslant N} \frac{1}{\|\mathrm{I} + \mathbf{m}\|^\alpha} < \begin{cases} s!c(\alpha, \varepsilon)^s N^{1+\varepsilon} \|\mathrm{I}\|^{-\alpha}, & if\ 1 \geqslant \alpha > 0, \\ s!c(\alpha)^s N^\alpha \|\mathrm{I}\|^{-\alpha}, & if\ \alpha > 1. \end{cases}$$

*Proof.* Suppose that $1 \geqslant \alpha > 0$. Since $N \leqslant \bar{l}_1/3$ and

$$\sum_{\bar{m}_1 \leqslant N} \frac{1}{(\overline{l_1 + m_1})^\alpha} \leqslant \left(\frac{3}{2}\right)^\alpha \frac{3N}{\bar{l}_1^\alpha}$$

for $s = 1$, the lemma is true for $s = 1$. Suppose that $k \geqslant 1$ and the lemma holds for $1 \leqslant s \leqslant k$. Now we proceed to prove that the lemma is also true for $s = k + 1$. Obviously, it follows from $\bar{m}_1 \cdots \bar{m}_{k+1} \leqslant \bar{l}_1 \cdots\cdots \bar{l}_{k+1}/3^{k+1}$ that there exists at least an $m_i$ such that $\bar{m}_i < \bar{l}_i/2$, where $1 \leqslant i \leqslant k + 1$. Hence

$$\sum_{\bar{m}_1 \cdots \bar{m}_{k+1} \leqslant N} \frac{1}{((\overline{l_1 + m_1}) \cdots (\overline{l_{k+1} + m_{k+1}}))^\alpha} \leqslant \Sigma_1 + \cdots + \Sigma_{k+1},$$

where

$$\Sigma_i = \sum_{\substack{\bar{m}_1 \cdots \bar{m}_{k+1} \leqslant N \\ \bar{m}_i \leqslant \bar{l}_i/2}} \frac{1}{((\overline{l_1 + m_1}) \cdots (\overline{l_{k+1} + m_{k+1}}))^\alpha}, \quad 1 \leqslant i \leqslant k + 1.$$

Suppose that $N \leqslant \bar{l}_2 \cdots \bar{l}_{k+1}/3^k$. Then by the induction hypothesis, we have

$$\begin{aligned} \Sigma_1 &\leqslant \frac{2^\alpha}{\bar{l}_1^\alpha} \sum_{\bar{m}_1 \leqslant N} \sum_{\bar{m}_2 \cdots \bar{m}_{k+1} \leqslant N/\bar{m}_1} \frac{1}{((\overline{l_2 + m_2}) \cdots (\overline{l_{k+1} + m_{k+1}}))^\alpha} \\ &\leqslant \frac{k!c(\alpha, \varepsilon)^k N^{1+\varepsilon}}{(\bar{l}_1 \cdots \bar{l}_{k+1})^\alpha} \sum_{\bar{m}_1 \leqslant N} \frac{1}{\bar{m}_1^{1+\varepsilon}} \leqslant \frac{k!c(\alpha, \varepsilon)^{k+1} N^{1+\varepsilon}}{(\bar{l}_1 \cdots \bar{l}_{k+1})^\alpha}. \end{aligned}$$

Suppose that $N > \bar{l}_2 \cdots \bar{l}_{k+1}/3^k$. Then

$$\Sigma_1 \leqslant \sigma_1 + \sigma_2,$$

where

$$\sigma_1 = \frac{2^\alpha}{\bar{l}_1^\alpha} \sum_{\bar{m}_1 \leqslant 3^k N/l_2 \cdots l_{k+1}} \sum_{\bar{m}_2 \cdots \bar{m}_{k+1} \leqslant N/\bar{m}_1} \frac{1}{((\overline{l_2 + m_2}) \cdots (\overline{l_{k+1} + m_{k+1}}))^\alpha}$$

and

$$\sigma_2 = \frac{2^\alpha}{\bar{l}_1^\alpha} \sum_{3^k N/l_2 \cdots l_{k+1} < \bar{m}_1 \leqslant N} \sum_{\bar{m}_2 \cdots \bar{m}_{k+1} \leqslant N/\bar{m}_1} \frac{1}{((\overline{l_2 + m_2}) \cdots (\overline{l_{k+1} + m_{k+1}}))^\alpha}.$$

Evidently

$$\begin{aligned}
\sigma_1 &\leqslant \frac{2^\alpha}{\bar{l}_1^\alpha} \sum_{\bar{m}_1 \leqslant 3^k N/l_2 \cdots l_{k+1}} \sum_{\bar{m}_2 \cdots \bar{m}_{k+1} \leqslant N/\bar{m}_1} \frac{(\bar{m}_2 \cdots \bar{m}_{k+1})^{1-\alpha+\varepsilon}}{(\bar{m}_2 \cdots \bar{m}_{k+1})^{1+\varepsilon}} \\
&\leqslant \frac{c(\alpha, \varepsilon)^k N^{1-\alpha+\varepsilon}}{\bar{l}_1^\alpha} \sum_{\bar{m}_1 \leqslant 3^k N/\bar{l}_2 \cdots \bar{l}_{k+1}} \frac{\bar{m}_1^\alpha}{\bar{m}_1^{1+\varepsilon}} \\
&\leqslant \frac{c(\alpha, \varepsilon)^k N^{1+\varepsilon}}{(\bar{l}_1 \cdots \bar{l}_{k+1})^\alpha} \sum \frac{1}{\bar{m}_1^{1+\varepsilon}} \leqslant \frac{c(\alpha, \varepsilon)^{k+1} N^{1+\varepsilon}}{(\bar{l}_1 \cdots \bar{l}_{k+1})^\alpha}
\end{aligned}$$

and by the induction hypothesis, we have

$$\sigma_2 \leqslant \frac{k! c(\alpha, \varepsilon)^k N^{1+\varepsilon}}{(\bar{l}_1 \cdots \bar{l}_{k+1})^\alpha} \sum \frac{1}{\bar{m}_1^{1+\varepsilon}} \leqslant \frac{k! c(\alpha, \varepsilon)^{k+1} N^{1+\varepsilon}}{(\bar{l}_1 \cdots \bar{l}_{k+1})^\alpha}.$$

Hence it follows that

$$\Sigma_1 \leqslant \frac{k! c(\alpha, \varepsilon)^{k+1} N^{1+\varepsilon}}{(\bar{l}_1 \cdots \bar{l}_{k+1})^\alpha}.$$

Since $\Sigma_i$ satisfies also the above inequality for $2 \leqslant i \leqslant k+1$, the lemma follows by mathematical induction. For the case $\alpha > 1$, the lemma may be proved similarly.

**Lemma 9.3** *Suppose that $\alpha$ and $Q$ are numbers satisfying $1 \geqslant \alpha > 0$ and $Q \geqslant 1$. If the congruence*

$$(\mathbf{a}, \mathbf{m}) = \sum_{k=1}^{s} a_k m_k \equiv 0 (\bmod\ n) \tag{9.3}$$

*has no solution in the domain*

$$\|\mathbf{m}\| \leqslant M, \quad \mathbf{m} \neq \mathbf{0}, \tag{9.4}$$

*then*

$$\sideset{}{'}\sum_{\substack{(\mathbf{a}, \mathbf{m}) \equiv 0 (\bmod n) \\ \|\mathbf{m}\| \leqslant Q}} \frac{1}{\|\mathbf{m}\|^\alpha} \leqslant c(\varepsilon)^s Q^{1-\alpha+\varepsilon} M^{-1}.$$

*Proof.* By Theorem 7.9, we have

$$\sum_{(\mathbf{a},\mathbf{m})\equiv 0 (\mathrm{mod}\ n)}{}' \frac{1}{\|\mathbf{m}\|^{1+\varepsilon}} \leqslant c(\varepsilon)^s M^{-1}.$$

Hence

$$\sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ n)\\ \|\mathbf{m}\|\leqslant Q}}{}' \frac{1}{\|\mathbf{m}\|^{\alpha}} \leqslant Q^{1-\alpha+\varepsilon} \sum_{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod}\ n)}{}' \frac{1}{\|\mathbf{m}\|^{1+\varepsilon}}$$
$$\leqslant c(\varepsilon)^s Q^{1-\alpha+\varepsilon} M^{-1}.$$

The lemma is proved.

**Lemma 9.4** *Suppose that* $n = p$ *and* $M = (4\zeta(1+\varepsilon)+2)^{-s} p^{1-\varepsilon}$. *Then there exists an integral vector* $\mathbf{a} = (1, a, \cdots a^{s-1})$ *such that the congruence* (9.3) *has no solution in the domain* (9.4).

*Proof.* Let $\mathbf{a} = (1, a, \cdots, a^{s-1})$. If $\mathbf{m} \neq 0$ and $\|\mathbf{m}\| \leqslant M$, then the congruence (9.3) has at most $s-1$ solutions in the range $1 \leqslant a \leqslant p$ (Cf. Lemma 4.5). Hence the total number of solutions of the congruence (9.3) in the domain $\|\mathbf{m}\| \leqslant M, \mathbf{m} \neq \mathbf{0}$ and $1 \leqslant a \leqslant p$ does not exceed

$$\sum_{\|\mathbf{m}\|\leqslant M}{}' \sum_{\substack{(\mathbf{a},\mathbf{m})\equiv 0(\mathrm{mod} p)\\ 1\leqslant a\leqslant p}} 1 \leqslant (s-1) \sum_{\|\mathbf{m}\|\leqslant M}{}' \frac{\|\mathbf{m}\|^{1+\varepsilon}}{\|\mathbf{m}\|^{1+\varepsilon}}$$
$$\leqslant (s-1)(2\zeta(1+\varepsilon)+1)^s M^{1+\varepsilon} < p/2.$$

Consequently, there exists an integer $a$ satisfying $1 \leqslant a \leqslant p$ such that the congruence (9.3) has no solution in the domain (9.4). The lemma is proved.

## 9.4 The approximate formula of the function of $E_s^\alpha\ (C)$

In this section, we suppose that $\alpha > 1$ and use the notations

$$\Delta_1 = \sup_{f\in E_s^\alpha(C)} \left\| \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) \sum_{\|\mathbf{m}\|<N} e^{2\pi i(\mathbf{m},\mathbf{x}-\frac{k\mathbf{a}}{n})} - f \right\|_2$$

and

$$\Delta_2 = \sup_{f\in E_s^\alpha(C)} \sup_{\mathbf{x}\in G_s} \left| \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) \sum_{\|\mathbf{m}\|<N} e^{2\pi i(\mathbf{m},\mathbf{x}-\frac{k\mathbf{a}}{n})} - f \right|.$$

**Theorem 9.4** *Suppose that* $N = [M^{\frac{2\alpha}{4\alpha-1}}]$. *If the congruence* (9.3) *has no solution in the domain* (9.4), *then*

$$\Delta_1 \leqslant Cs!^{1/2} c(\alpha,\varepsilon)^s M^{-\frac{\alpha(2\alpha-1)}{4\alpha-1}+\varepsilon}.$$

*Proof.* Clearly, we have

$$\Delta_1^2 \leqslant \Sigma_1 + \Sigma_2, \tag{9.5}$$

where

$$\Sigma_1 = \sup_{f\in E_s^\alpha(C)} \sum_{\|\mathbf{m}\|<N} \left| C(\mathbf{m}) - \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) e^{-2\pi i(\mathbf{m},\frac{k\mathbf{a}}{n})}\right|^2$$

and

$$\Sigma_2 = \sup_{f\in E_s^\alpha(C)} \sum_{\|\mathbf{m}\|\geqslant N} |C(\mathbf{m})|^2.$$

It follows by Lemma 3.6 that

$$\frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) e^{-2\pi i(\mathbf{m},\frac{k\mathbf{a}}{n})} = \frac{1}{n}\sum_{k=1}^{n}\sum C(\mathbf{l}) e^{2\pi i(l-m,a)k/n}$$

$$= \sum_{(\mathbf{a},\mathbf{l}-\mathbf{m})\equiv 0(\mathrm{mod} n)} C(\mathbf{l}) = \sum_{(\mathbf{a},\mathbf{r})\equiv 0(\mathrm{mod} n)} C(\mathbf{r}+\mathbf{m}),$$

$$\left| C(\mathbf{m}) - \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) e^{-2\pi i(\mathbf{m},\frac{k\mathbf{a}}{n})}\right| \leqslant C \sideset{}{'}\sum_{(\mathbf{a},\mathbf{r})\equiv 0(\mathrm{mod}\ n)} \frac{1}{\|\mathbf{r}+\mathbf{m}\|^\alpha}.$$

We may suppose that $N \leqslant M/3^s$. Since

$$\|\mathbf{r}\| \leqslant 2^s\|\mathbf{m}\|\|\mathbf{r}+\mathbf{m}\|,$$

therefore by Theorem 7.9 and Lemma 9.2, we have

$$\begin{aligned}
\Sigma_1 &\leqslant C^2 \sum_{\|\mathbf{m}\|<N}\left(\sideset{}{'}\sum_{(\mathbf{a},\mathbf{r})\equiv 0(\mathrm{mod}\ n)} \frac{1}{\|\mathbf{r}+\mathbf{m}\|^\alpha}\right)^2 \\
&= C^2 \sum_{\|\mathbf{m}\|<N}\ \sideset{}{'}\sum_{(\mathbf{a},\mathbf{r})\equiv 0(\mathrm{mod}\ n)} \frac{1}{\|\mathbf{r}+\mathbf{m}\|^\alpha} \sideset{}{'}\sum_{(\mathbf{a},\mathbf{r}')\equiv 0(\mathrm{mod}\ n)} \frac{\|\mathbf{m}\|^\alpha}{\|\mathbf{r}'\|^\alpha}\left(\frac{\|\mathbf{r}'\|}{\|\mathbf{m}\|\|\mathbf{r}'+\mathbf{m}\|}\right)^\alpha \\
&\leqslant C^2 2^{\alpha s} N^\alpha \sideset{}{'}\sum_{(\mathbf{a},\mathbf{r}')\equiv 0(\mathrm{mod}\ n)} \frac{1}{\|\mathbf{r}'\|^2} \sideset{}{'}\sum_{(\mathbf{a},\mathbf{r})\equiv 0(\mathrm{mod}\ n)} \sum_{\|\mathbf{m}\|<N} \frac{1}{\|\mathbf{r}+\mathbf{m}\|^\alpha} \\
&\leqslant C^2 s! c(\alpha)^s N^{2\alpha}\left(\sideset{}{'}\sum_{(\mathbf{a},\mathbf{r})\equiv 0(\mathrm{mod}\ n)} \frac{1}{\|\mathbf{r}\|^\alpha}\right)^2 \\
&\leqslant C^2 s! c(\alpha,\varepsilon)^s N^{2\alpha} M^{-2\alpha+\varepsilon} \leqslant C^2 s! c(\alpha,\varepsilon)^s M^{-\frac{2\alpha(2\alpha-1)}{4\alpha-1}+\varepsilon}
\end{aligned} \tag{9.6}$$

The $\Sigma_2$ may be estimated as follows.

$$\Sigma_2 \leqslant C^2 \sum_{\|\mathbf{m}\|\geqslant N} \frac{1}{\|\mathbf{m}\|^{2\alpha}} = C^2 \sum_{\|\mathbf{m}\|\geqslant N} \frac{\|\mathbf{m}\|^{-2\alpha+1+\varepsilon}}{\|\mathbf{m}\|^{1+\varepsilon}}$$

$$= C^2 N^{-2\alpha+1+\varepsilon} \sum \frac{1}{\|\mathbf{m}\|^{1+\varepsilon}} = C^2 c(\varepsilon)^s N^{-2\alpha+1+\epsilon}$$
$$= C^2 c(\varepsilon)^s M^{-\frac{2\alpha(2\alpha-1)}{4\alpha-1}+\varepsilon}. \tag{9.7}$$

The theorem follows by (9.5), (9.6) and (9.7).

**Theorem 9.5** *Suppose that* $N = [M^{\frac{\alpha}{2\alpha-1}}]$. *If the congruence* (9.3) *has no solution in* (9.4), *then*

$$\Delta_2 \leqslant Cs!c(\alpha,\varepsilon)^s M^{-\frac{\alpha(\alpha-1)}{2\alpha-1}+\varepsilon}.$$

*Proof.* Evidently

$$\Delta_2 \leqslant \Sigma_1 + \Sigma_2,$$

where

$$\Sigma_1 = \sup_{f\in E_s^\alpha(C)} \sum_{\|\mathbf{m}\|<N} \left| C(\mathbf{m}) - \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) e^{-2\pi i(\mathbf{m},\frac{k\mathbf{a}}{n})} \right|$$

and

$$\Sigma_2 = \sup_{f\in E_s^\alpha(C)} \sum_{\|\mathbf{m}\|\geqslant N} |C(\mathbf{m})|.$$

Similar to (9.6) and (9.7), we have

$$\Sigma_1 \leqslant Cs!c(\alpha,\varepsilon)^s N^\alpha M^{-\alpha+\varepsilon} = Cs!c(\alpha,\varepsilon)^s M^{-\frac{\alpha(\alpha-1)}{2\alpha-1}+\varepsilon}$$

and

$$\Sigma_2 \leqslant C \sum_{\|\mathbf{m}\|\geqslant N} \frac{1}{\|\mathbf{m}\|^\alpha} \leqslant Cc(\varepsilon)^s N^{-\alpha+1+\varepsilon}$$
$$\leqslant Cc(\varepsilon)^s M^{-\frac{\alpha(\alpha-1)}{2\alpha-1}+\varepsilon}.$$

The theorem follows.

By Theorem 9.4 and Lemma 7.5 (with the notation of §4.6), we have

**Theorem 9.6** *Suppose that* $s = \dfrac{\varphi(m)}{2}$ *and* $\mathbf{a} = (c_1, \cdots, c_s)$. *Then*

$$\Delta_1 \leqslant Cc(\mathscr{R}_s,\alpha,\varepsilon) n^{-(\frac{1}{2}+\frac{1}{2(s-1)})\frac{\alpha(2\alpha-1)}{4\alpha-1}+\varepsilon}.$$

From Theorem 9.4 and Lemma 9.4, we have:

**Theorem 9.7** *Suppose that* $n = p$. *Then there exists an integral vector* $\mathbf{a}$ $(= \mathbf{a}(p))$ *such that*

$$\Delta_1 \leqslant Cs!^{1/2} c(\alpha,\varepsilon)^s p^{-\frac{\alpha(2\alpha-1)}{4\alpha-1}+\varepsilon}.$$

## 9.5 The approximate formula of the function of $Q_s^\alpha(C)$

Introduce the notations

$$\mu(\alpha) = \begin{cases} \dfrac{\alpha}{2}, & \text{if } \alpha > 1, \\ \dfrac{2\alpha^2}{1+4\alpha-\alpha^2}, & \text{if } 1 \geqslant \alpha > 0 \end{cases}$$

and

$$\Delta = \sup_{f \in Q_s^\alpha(C)} \left\| \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) \sum_{\|\mathbf{m}\|<N} e^{2\pi i(\mathbf{m},\mathbf{x}-\frac{k\mathbf{a}}{n})} - f \right\|_2 .$$

**Theorem 9.8** *Suppose that* $N = [M^{\frac{\mu(\alpha)}{\alpha}}]$. *If the congruence* (9.3) *has no solution in* (9.4), *then*

$$\Delta \leqslant Cs!^{1/2} c(\alpha,\varepsilon)^s M^{-\mu(\alpha)+\varepsilon}.$$

*Proof.* Let

$$T = \begin{cases} [\log_2 M] + 1, & \text{if } \alpha > 1, \\ [\log_2 MN^{-1+\alpha/2}] + 1, & \text{if } 1 \geqslant \alpha > 0. \end{cases}$$

Then by Minkowski's inequality, we have

$$\Delta \leqslant \Sigma_1 + \Sigma_2 + \Sigma_3,$$

where

$$\Sigma_1 = \sup_{f \in Q_s^\alpha(C)} \left\| f(\mathbf{x}) - \sum_{t_0 \leqslant T}{}'' \varphi_{\mathbf{t}}(\mathbf{x}) \right\|_2 ,$$

$$\Sigma_2 = \sup_{f \in Q_s^\alpha(C)} \left\| \sum_{t_0 \leqslant T}{}'' \left( \varphi_{\mathbf{t}}(\mathbf{x}) - \frac{1}{n} \sum_{k=1}^{n} \varphi_{\mathbf{t}}\left(\frac{k\mathbf{a}}{n}\right) \sum_{\|\mathbf{m}\|<N} e^{2\pi i(\mathbf{m},\mathbf{x}-\frac{k\mathbf{a}}{n})} \right) \right\|_2$$

and

$$\Sigma_3 = \sup_{f \in Q_s^\alpha(C)} \left\| \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) \sum_{\|\mathbf{m}\|<N} e^{2\pi i(\mathbf{m},\mathbf{x},-\frac{k\mathbf{a}}{n})} - \sum_{t_0 \leqslant T}{}'' \frac{1}{n} \sum_{k=1}^{n} \varphi_{\mathbf{t}}\left(\frac{k\mathbf{a}}{n}\right) \sum_{\|\mathbf{m}\|<N} e^{2\pi i(\mathbf{m},\mathbf{x}-\frac{k\mathbf{a}}{n})} \right\|_2 .$$

It follows from Minkowski's inequality that

$$\Sigma_1 \leqslant \sup_{f \in Q_s^\alpha(C)} \sum_{t_0 > T}{}'' \|\varphi_{\mathbf{t}}\|_2 \leqslant C \sum_{t_0 > T}{}'' 2^{-(\alpha-\varepsilon)t_0 - \varepsilon t_0}$$

$$\leqslant Cc(\alpha,\varepsilon)^s 2^{-(\alpha-\varepsilon)T} \leqslant Cc(\alpha,\varepsilon)^s M^{-\mu(\alpha)+\varepsilon}.$$

Evidently

$$\Sigma_2 \leqslant \sigma_1 + \sigma_2,$$

where

$$\sigma_1 = \sup_{f\in Q_s^\alpha(C)} \left\| \sum_{t_0\leqslant T}'' \sum_{\|\mathbf{m}\|<N} \left( C_{\mathbf{t}}(\mathbf{m}) - \frac{1}{n}\sum_{k=1}^{n} \varphi_{\mathbf{t}}\left(\frac{k\mathbf{a}}{n}\right) e^{-2\pi i(\mathbf{m},\frac{k\mathbf{a}}{n})} \right) e^{2\pi i(\mathbf{m},\mathbf{x})} \right\|_2$$

and

$$\sigma_2 = \sup_{f\in Q_s^\alpha(C)} \left\| \sum_{t_0\leqslant T}'' \sum_{\|\mathbf{m}\|\geqslant N} C_{\mathbf{t}}(\mathbf{m}) e^{2\pi i(\mathbf{m},\mathbf{x})} \right\|_2 .$$

Since

$$C_{\mathbf{t}}(\mathbf{m}) - \frac{1}{n}\sum_{k=1}^{n} \varphi_{\mathbf{t}}\left(\frac{k\mathbf{a}}{n}\right) e^{-2\pi i(\mathbf{m},\frac{k\mathbf{a}}{n})} = -\sum_{(\mathbf{a},\mathbf{l})\equiv 0(\bmod n)}' C_{\mathbf{t}}(\mathbf{l}+\mathbf{m})$$

and

$$\|\mathbf{l}\| \leqslant 2^s \|\mathbf{m}\| \|\mathbf{l}+\mathbf{m}\|,$$

it follows by Theorem 6.2 that

$$\begin{aligned}
\sigma_1^2 &\leqslant \sup_{f\in Q_s^\alpha(C)} \left\| \sum_{t_0\leqslant T}'' \sum_{\|\mathbf{m}\|<N} \sum_{(\mathbf{a},\mathbf{l})\equiv 0(\bmod n)}' C_{\mathbf{t}}(\mathbf{l}+\mathbf{m}) e^{2\pi i(\mathbf{m},\mathbf{x})} \right\|_2^2 \\
&\leqslant \sup_{f\in Q_s^\alpha(C)} \sum_{\|\mathbf{m}\|<N} \left( \sum_{t_0\leqslant T}'' \sum_{(\mathbf{a},\mathbf{l})\equiv 0(\bmod\ n)}' |C_{\mathbf{t}}(\mathbf{l}+\mathbf{m})| \right)^2 \\
&\leqslant C^2 c(\alpha)^s \sum_{\|\mathbf{m}\|<N} \left( \sum_{\substack{(\mathbf{a},\mathbf{l})\equiv 0(\bmod\ n)\\ \|\mathbf{l}\|\leqslant 2^{s+T}N}}' \frac{1}{\|\mathbf{l}+\mathbf{m}\|^\alpha} \right)^2 \\
&\leqslant C^2 c(\alpha)^s N^\alpha \sum_{\|\mathbf{m}\|<N} \sum_{\substack{(\mathbf{a},\mathbf{l})\equiv 0(\bmod\ n)\\ \|\mathbf{l}\|\leqslant 2^{s+T}N}}' \frac{1}{\|\mathbf{l}+\mathbf{m}\|^\alpha} \sum_{\substack{(\mathbf{a},\mathbf{l}')\equiv 0(\bmod\ n)\\ \|\mathbf{l}'\|\leqslant 2^{s+T}N}}' \frac{1}{\|\mathbf{l}'\|^\alpha}.
\end{aligned}$$

We may suppose that $N \leqslant M/3^s$. Hence by Theorem 7.9, Lemmas 9.2 and 9.3, we have

$$\sigma_1^2 \leqslant \begin{cases} C^2 s! c(\alpha,\varepsilon)^s N^{2\alpha} M^{-2\alpha+\varepsilon}, & \text{if } \alpha > 1, \\ C^2 s! c(\alpha,\varepsilon)^s N^{3-\alpha+\varepsilon} M^{-2} 2^{2T(1-\alpha)+T\varepsilon}, & \text{if } 1 \geqslant \alpha > 0. \end{cases}$$

By Schwarz's inquality,

$$\begin{aligned}
\sigma_2^2 &\leqslant \sup_{f\in Q_s^\alpha(C)} \sum_{\|\mathbf{m}\|\geqslant N} \left( {\sum_{t_0\leqslant T}}'' |C_{\mathbf{t}}(\mathbf{m})| \right)^2 \\
&\leqslant \sup_{f\in Q_s^\alpha(C)} \sum_{\|\mathbf{m}\|\geqslant N} {\sum_{t_0\leqslant T}}'' 2^{-\varepsilon t_0/2} {\sum_{t_0\leqslant T}}'' 2^{\varepsilon t_0/2} |C_{\mathbf{t}}(\mathbf{m})|^2 \\
&\leqslant c(\varepsilon)^s \sup_{f\in Q_s^\alpha(C)} {\sum_{t_0\leqslant T}}'' \sum_{\|\mathbf{m}\|\geqslant N} 2^{\varepsilon t_0/2} |C_{\mathbf{t}}(\mathbf{m})|^2 \\
&\leqslant c(\varepsilon)^s \sup_{f\in Q_s^\alpha(C)} {\sum_{t_0>\log_2 N}}'' 2^{\varepsilon t_0/2} \|\varphi_{\mathbf{t}}\|_2^2 \\
&\leqslant C^2 c(\varepsilon)^s {\sum_{t_0>\log_2 N}}'' 2^{-2\alpha t_0+\varepsilon t_0/2} \\
&\leqslant C^2 c(\alpha,\varepsilon)^s N^{-2\alpha+\varepsilon}.
\end{aligned}$$

Hence

$$\Sigma_2 \leqslant Cs!^{1/2} c(\alpha,\varepsilon)^s M^{-\mu(\alpha)+\varepsilon}.$$

$$\begin{aligned}
\Sigma_3 &\leqslant \sup_{f\in Q_s^\alpha(C)} \left\| {\sum_{t_0>T}}'' \frac{1}{n} \sum_{k=1}^{n} \varphi_{\mathbf{t}}\left(\frac{k\mathbf{a}}{n}\right) \sum_{\|\mathbf{m}\|<N} e^{2\pi i(\mathbf{m},\mathbf{x}-\frac{k\mathbf{a}}{m})} \right\|_2 \\
&\leqslant \sup_{f\in Q_s^\alpha(C)} {\sum_{t_0>T}}'' \|\varphi_{\mathbf{t}}\| \left( \sum_{\|\mathbf{m}\|<N} 1 \right)^{1/2} \\
&\leqslant C {\sum_{t_0>T}}'' 2^{-\alpha t_0} \left( \sum_{\|\mathbf{m}\|<N} \frac{N^{1+\varepsilon}}{\|\mathbf{m}\|^{1+\varepsilon}} \right)^{1/2} \\
&\leqslant Cc(\alpha,\varepsilon)^s N^{\frac{1}{2}+\frac{\varepsilon}{2}} 2^{-\alpha T+\frac{\varepsilon T}{2}} \\
&\leqslant Cc(\alpha,\varepsilon)^s M^{-\mu(\alpha)+\varepsilon}.
\end{aligned}$$

The theorem follows.

From Theorem 9.8 and Lemma 7.5, we have

**Theorem 9.9** *Suppose that* $s = \dfrac{\varphi(m)}{2}$ *and* $\mathbf{a} = (c_1, \cdots, c_s)$. *Then*

$$\Delta \leqslant Cc(\mathscr{R}_s, \alpha, \varepsilon) n^{-(\frac{1}{2}+\frac{1}{2(s-1)})\mu(\alpha)+\varepsilon}.$$

From Theorem 9.8 and Lemma 9.4, we have

**Theorem 9.10** *Suppose that* $n = p$. *Then there exists an integral vector* $\mathbf{a}(= \mathbf{a}(p))$ *such that*

$$\Delta \leqslant Cs!^{1/2} c(\alpha,\varepsilon)^s p^{-\mu(\alpha)+\varepsilon}.$$

## 9.6 The Bernoulli polynomial and the approximate polynomial

Suppose that $\alpha > 1$. We shall use the Bernoulli polynomials to express the function $f$ of $E_s^\alpha(C)$ as a definite integral over $G_s$ and then by the use of the results of numerical integration to obtain the approximate polynomial of $f$. The resuts so obtained are sharper than those given in §9.4 for certain values of $\alpha$.

**Lemma 9.5** *If $f_i \in E_s^\alpha(C_i)(i = 1, 2)$, then $f_1 f_2 \in E_s^\alpha(C_1 C_2 c(\alpha)^s)$.*

*Proof.* Let $C_i(\mathbf{m})$ and $C(\mathbf{m})$ be the Fourier coefficients of $f_i(i = 1, 2)$ and $f$ respectively. Then

$$\begin{aligned} f(\mathbf{x}) &= \sum_{\mathbf{n}} \sum_{\mathbf{k}} C_1(\mathbf{n}) C_2(\mathbf{k}) e^{2\pi i(\mathbf{n}+\mathbf{k},\mathbf{x})} \\ &= \sum_{\mathbf{m}} \left( \sum_{\mathbf{n}} C_1(\mathbf{n}) C_2(\mathbf{m}-\mathbf{n}) \right) e^{2\pi i(\mathbf{m},\mathbf{x})} \\ &= \sum_{\mathbf{m}} C(\mathbf{m}) e^{2\pi i(\mathbf{m},\mathbf{x})}, \end{aligned}$$

where

$$C(\mathbf{m}) = \sum_{\mathbf{n}} C_1(\mathbf{n}) C_2(\mathbf{m}-\mathbf{n}).$$

Since

$$\begin{aligned} \sum_{n=-\infty}^{\infty} \frac{1}{(\bar{n}(\overline{m-n}))^\alpha} &= \sum_{|n| \leqslant |m|/2} \frac{1}{(\bar{n}(\overline{m-n}))^\alpha} + \sum_{|n| > |m|/2} \frac{1}{(\bar{n}(\overline{m-n}))^\alpha} \\ &\leqslant \frac{2^\alpha}{\bar{m}^\alpha} \left( \sum_{|n| \leqslant |m|/2} \frac{1}{\bar{n}^\alpha} + \sum_{|n| > |m|/2} \frac{1}{(\overline{m-n})^\alpha} \right) \\ &\leqslant \frac{2^{\alpha+1}}{\bar{m}^\alpha} \sum \frac{1}{\bar{n}^\alpha} = c(\alpha) \bar{m}^{-\alpha}, \end{aligned}$$

therefore

$$\begin{aligned} |C(\mathbf{m})| &\leqslant \left| \sum_{\mathbf{n}} C_1(\mathbf{n}) C_2(\mathbf{m}-\mathbf{n}) \right| \\ &\leqslant C_1 C_2 \sum_{\mathbf{n}} \frac{1}{(\|\mathbf{n}\| \|\mathbf{m}-\mathbf{n}\|)^\alpha} \\ &= C_1 C_2 \prod_{i=1}^{s} \left( \sum_{n_i=-\infty}^{\infty} \frac{1}{(\bar{n}_i(\overline{m_i-n_i}))^\alpha} \right) \\ &= C_1 C_2 c(\alpha)^s \|\mathbf{m}\|^{-\alpha}. \end{aligned}$$

The lemma is proved.

**Lemma 9.6** *Suppose that $f$ has $r$-th continuous derivatives and $f \in E_1^\alpha(C)$. Then.*

$$f(x) = \int_0^1 \sum_{\tau=0}^{1} f^{(v\tau)}(y)\varphi_v^\tau(y-x)dy$$

*for $v = 1, \cdots, r$, where*

$$\varphi_v(x) = \frac{(-1)^{v-1}}{v!} B_v(\{x\})$$

*in which $B_v(x)$ denotes the $v$-th Bernoulli polynomial.*

*Proof.* Since

$$\begin{aligned}
&\int_0^1 \sum_{\tau=0}^{1} f^{(v\tau)}(y)\varphi_v^\tau(y-x)dy \\
&= \int_0^1 f(y)dy + \frac{(-1)^{v-1}}{v!}\int_0^1 f^{(v)}(y)B_v(\{y-x\})dy \\
&= \int_0^1 f(y)dy + \frac{(-1)^{v-1}}{v!}\int_0^1 f^{(v)}(x+y)B_v(y)dy,
\end{aligned}$$

the lemma is reduced to proving by induction the assertion that the right hand is equal to $f(x)$. Since

$$\begin{aligned}
\int_0^1 f'(x+y)B_1(y)dy &= f(x+y)B_1(y)\Big|_0^1 - \int_0^1 f(x+y)dy \\
&= f(x) - \int_0^1 f(y)dy,
\end{aligned}$$

the assertion holds for $v = 1$. Suppose that $r \geqslant v \geqslant 2$ and the assertion holds for any positive integer less than $v$. Since

$$B_v(1) = B_v(0), \quad B_v'(y) = vB_{v-1}(y)$$

by Lemma 6.6, therefore

$$\begin{aligned}
&\frac{(-1)^{v-1}}{v!}\int_0^1 f^{(v)}(x+y)B_v(y)dy \\
&= \frac{(-1)^{v-1}}{v!} f^{(v-1)}(x+y)B_v(y)\Big|_0^1 \\
&\quad - \frac{(-1)^{v-1}}{v!}\int_0^1 f^{(v-1)}(x+y)B_v'(y)dy \\
&= \frac{(-1)^{v-2}}{(v-1)!}\int_0^1 f^{(v-1)}(x+y)B_{v-1}(y)dy
\end{aligned}$$

$$= \cdots = f(x) - \int_0^1 f(y)dy.$$

Hence the assertion holds for $v$. The lemma follows.

**Lemma 9.7** *Suppose that $f$ has continuous derivatives $f(\mathbf{x})^{(r_1,\cdots,r_s)}(0 \leqslant r_1,\cdots, r_s \leqslant r)$ and $f \in E_s^\alpha(C)$. Then*

$$f(\mathbf{x}) = \int_{G_s} \sum_{\tau_1,\cdots,\tau_s=0}^{1} f(\mathbf{y})^{(\tau_1 v,\cdots,\tau_s v)} \prod_{i=1}^{s} \varphi_v^{\tau_i}(y_i - x_i)d\mathbf{y}$$

*holds for $v = 1, \cdots, r$.*

*Proof.* The lemma is true for $s = 1$ by Lemma 9.6. Now suppose that $s \geqslant 2$ and the lemma holds for any positive integer less than $s$. Obviously $f \in E_{s-1}^\alpha(Cc(\alpha))$. Hence it follows by the induction hypothesis that

$$f(\mathbf{x}) = \int_{G_{s-1}} \sum_{\tau_1,\cdots,\tau_{s-1}=0}^{1} f(y_1,\cdots,y_{s-1},x_s)^{(\tau_1 v,\cdots,\tau_{s-1}v,0)} \cdot \prod_{i=1}^{s-1} \varphi_v^{\tau_v}(y_v - x_v)dy_1\cdots dy_{s-1} \tag{9.8}$$

holds for $v = 1, \cdots, r$. Since

$$f(y_1,\cdots,y_{s-1},x_s)^{(\tau_1 v,\cdots,\tau_{s-1}v,0)} = \int_0^1 \sum_{\tau_s=0}^{1} f(\mathbf{y})^{(\tau_1 v,\cdots,\tau_s v)} \varphi_v^{\tau_s}(y_s - x_s)dy_s \tag{9.9}$$

by Lemma 9.6, the lemma follows by substituting (9.9) into (9.8).

Introduce the notation

$$\Delta = \sup_{f\in E_s^\alpha(C)} \sup_{x\in G_s} \left| \frac{1}{n}\sum_{k=1}^{n} \sum_{\tau_1,\cdots,\tau_s=0}^{1} f\left(\frac{k\mathbf{a}}{n}\right)^{(\tau_1 r,\cdots,\tau_s r)} \cdot \prod_{i=1}^{s} \varphi_r^{\tau_i}\left(\frac{a_i k}{n} - x_i\right) - f \right|.$$

**Theorem 9.11** *Suppose that $\alpha > 3$. If the congruence* (9.3) *has no solutions in* (9.4), *then*

$$\Delta \leqslant Cc(\alpha,\varepsilon)^s M^{-\gamma+\varepsilon},$$

*where $\gamma = \min(r, \alpha - r)$ and $r = \left[\frac{\alpha+1}{2}\right]$.*

*Proof.* Suppose that $f \in E_s^\alpha(C)$. Then

$$f(\mathbf{x})^{(\tau_1 r, \cdots, \tau_s r)} = (2\pi i)^{(\tau_1+\cdots+\tau_s)r} \sum \left(\prod_{i=1}^{s} m_i^{\tau_i r}\right) C(\mathbf{m}) e^{2\pi i(\mathbf{m},\mathbf{x})},$$

where $\tau_i = 0$ or $1 (1 \leqslant i \leqslant s)$. Since

$$\left| \left(\prod_{i=1}^{s} m_i^{\tau_i r}\right) C(\mathbf{m}) \right| \leqslant C\|\mathbf{m}\|^{-\alpha} \prod_{i=1}^{s} |m_i|^{\tau_i r} \leqslant C\|\mathbf{m}\|^{-\alpha+r},$$

we have $f(\mathbf{x})^{\tau_1 r, \cdots, \tau_s r} \in E_s^{\alpha-r}(C(2\pi)^{rs})$, where

$$\alpha - r \geqslant \alpha - \frac{\alpha+1}{2} = \frac{\alpha-1}{2} > 1.$$

And so it follows from Lemma 9.7 that

$$f(\mathbf{x}) = \sum_{\tau_1, \cdots, \tau_s = 0}^{1} \int_{G_s} f(\mathbf{y})^{(\tau_1 r, \cdots, \tau_s r)} \prod_{i=1}^{s} \varphi_r^{\tau_i}(y_i - x_i) d\boldsymbol{y}. \tag{9.10}$$

Let $C(k)$ be the Fourier coefficient of $B_r(\{x\})$. Then $B_r^{(t)}(0) = B_r^{(t)}(1)$ for $t \leqslant r-2$ by Lemma 6.7 and so

$$\begin{aligned} C(k) &= \int_0^1 B_r(x) e^{-2\pi i k x} dx \\ &= \frac{-1}{(2\pi i k)^r} B_r^{(r-1)}(x) e^{-2\pi i k x} \Big|_0^1 + \frac{1}{(2\pi i k)^r} \int_0^1 B_r^{(r)}(x) e^{-2\pi i k x} dx \\ &= c(r)|k|^{-r} \end{aligned}$$

for $k \neq 0$, i.e., $B_r(\{x\}) \in E_1^r(c(r))$. Since for any given $\mathbf{x}$,

$$\prod_{i=1}^{s} \varphi_r^{\tau_i}(y_i - x_i) = \frac{(-1)^{(\tau_1+\cdots+\tau_s)(r-1)}}{r!^{\tau_1+\cdots+\tau_s}} \prod_{i=1}^{s} B_r^{\tau_i}(\{y_i - x_i\}) \in E_s^r(c(r)^s),$$

we have

$$f(\mathbf{y})^{(\tau_1 r, \cdots, \tau_s r)} \prod_{i=1}^{s} \varphi_r^{\tau_i}(y_i - x_i) \in E_s^\gamma(Cc(\alpha)^s)$$

by Lemma 9.5. The theorem follows from (9.9) and Theorem 7.9.

For the case $\alpha = 2k (k = 2, 3, \cdots)$, we have $r = \alpha/2$ and $\gamma = \alpha/2$. Hence Theorem 9.11 is sharper than Theorems 9.4 and 9.5.

From Theorem 9.11 and Lemma 7.5, we have

**Theorem 9.12** *Suppose that* $s = \frac{\varphi(m)}{2}$ *and* $\mathbf{a} = (c_1, \cdots, c_s)$. *Then*

$$\Delta \leqslant Cc(\mathscr{R}_s, \alpha, \varepsilon) n^{-(\frac{1}{2}+\frac{1}{2(s-1)})\gamma+\varepsilon}.$$

By Theorem 9.11 and Lemma 9.4, we have

**Theorem 9.13** *Suppose that* $n = p$. *Then there exists an integral vector* $\mathbf{a} = \mathbf{a}(p)$ *such that*

$$\Delta \leqslant Cc(\alpha, \varepsilon)^s p^{-\gamma+\varepsilon}.$$

## 9.7 The $\Omega$ results

**Lemma 9.8** *Suppose that* $\xi$ *is a real number and* $p_k/q_k$ *is the* $k$-th *convergent of* $\xi$. *Then*

$$\left|\frac{p_k}{q_k} - \xi\right| \leqslant \frac{1}{q_k q_{k+1}}.$$

(Cf. Hua Loo Keng [2], Chap. 10).

**Theorem 9.14** *For any given integer* $N$ *satisfying* $n > N \geqslant 1$ *and any integral vector* $\mathbf{a}$, *we have*

$$\begin{aligned}\Delta &= \sup_{f \in H_s^\alpha(C)} \left\| \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) \sum_{\|\mathbf{m}\| < N} e^{-2\pi i(\mathbf{m}, \frac{k\mathbf{a}}{n} - \mathbf{x})} - f \right\|_2 \\ &\geqslant \frac{C}{2(2\pi)^{2\alpha}} n^{-\alpha/2}.\end{aligned}$$

*Proof.* Evidently, we may suppose that $N > 1$. Otherwise we have $\Delta \geqslant C$ for $f(\mathbf{x}) = C$. We may assume also that $(a_1, \cdots, a_s, n) = 1$ and $a_1 = -1$. Let $p_t/q_t$ be the $t$-th convergent of $\frac{p_h}{q_h} = \frac{a_2}{n}$ and $N$ satisfy

$$1 = q_0 < \cdots < q_k \leqslant N < q_{k+1} < \cdots < q_h \leqslant n.$$

Let

$$K = a_2 q_k - n p_k.$$

Then

$$|K| \leqslant \frac{n}{q_{k+1}} < \frac{n}{N}$$

by Lemma 9.8. Take

$$f(\mathbf{x}) = \frac{C}{4(2\pi)^{2\alpha}} \left( \frac{e^{2\pi i(Kx_1+x_2)}}{\bar{K}^\alpha} + \frac{e^{-2\pi i(Kx_1+x_2)}}{\bar{K}^\alpha} + \frac{e^{2\pi i N x_1}}{N^\alpha} + \frac{e^{-2\pi i N x_1}}{N^\alpha} \right).$$

Then

$$\begin{aligned}\Delta^2 &\geqslant \sum_{\bar{m}_1\bar{m}_2<N}\left|\sum_{l_1\equiv a_2l_2(\mathrm{mod}n)}{}' C(l_1+m_1,l_2+m_2,0,\cdots,0)\right|^2+2\frac{C^2}{16(2\pi)^{4\alpha}}N^{-2\alpha}\\ &\geqslant \sum_{\bar{m}_1\bar{m}_2<K}(C(K+m_1,q_k+m_2,0,\cdots,0)\\ &\quad +C(-K+m_1,-q_k+m_2,0,\cdots,0))^2+\frac{C^2}{8(2\pi)^{4\alpha}}N^{-2\alpha}\\ &\geqslant C(K,1,0,\cdots,0)^2+C(-K,-1,0,\cdots,0)^2+\frac{C^2}{8(2\pi)^{4\alpha}}N^{-2\alpha}\\ &\geqslant \frac{C^2}{8(2\pi)^{4\alpha}}(n^{-2\alpha}N^{2\alpha}+N^{-2\alpha})\geqslant\frac{C^2}{4(2\pi)^{4\alpha}}n^{-\alpha}.\end{aligned}$$

The theorem is proved.

**Theorem 9.15** *For any given functions $\psi_{n,k}(\mathbf{x})(1\leqslant k\leqslant n)$ and integral vector* $\mathbf{a}$, *we have*

$$\sup_{f\in H_s^\alpha(C)}\sup_{\mathbf{x}\in G_s}\left|\sum_{k=1}^n f\left(\frac{k\mathbf{a}}{n}\right)\psi_{n,k}(\mathbf{x})-f\right|\geqslant\frac{C}{(2\pi)^\alpha}n^{-\alpha/2}.$$

To prove Theorem 9.15, we shall need

**Lemma 9.9** *The congruence*

$$a_1n_1+a_2n_2\equiv 0(\mathrm{mod}n) \tag{9.11}$$

*has solution satisfying*

$$|n_1|\leqslant\sqrt{n},\quad |n_2|\leqslant\sqrt{n},\quad (n_1,n_2)\neq(0,0). \tag{9.12}$$

*Proof.* Since

$$([\sqrt{n}]+1)([\sqrt{n}]+1)>n,$$

the number of integer pairs $(x_1,x_2)(x_1,x_2=0,1,\cdots,[\sqrt{n}])$ is greater than $n$. Hence there exist two different pairs $(x_1',x_2')$ and $(x_1'',x_2'')$ such that

$$a_1x_1'+a_2x_2'\equiv a_1x_1''+a_2x_2''(\mathrm{mod}\ n).$$

Take $n_1=x_1'-x_1''$ and $n_2=x_2'-x_2''$. Then $n_1$ and $n_2$ satisfy (9.11) and (9.12). The lemma is proved.

The proof of Theorem 9.15. Let $n_1,n_2$ satisfy (9.11) and (9.12). Without loss of generality, we may suppose that $n_2\neq 0$. Then we have

$$a_1n_1k\equiv -a_2n_2k(\mathrm{mod}n) \tag{9.13}$$

for any integer $k$. Take

$$f(\mathbf{x}) = C\frac{e^{2\pi i n_1 x_1} - e^{-2\pi i n_2 x_2}}{2(2\pi)^\alpha N^\alpha} \in H_s^\alpha(C),$$

where $N = \max(|n_1|, |n_2|)$. Then

$$f\left(\frac{k\mathbf{a}}{n}\right) = C\frac{e^{\frac{2\pi i a_1 n_1 k}{n}} - e^{\frac{2\pi i a_2 n_2 k}{n}}}{2(2\pi)^\alpha N^\alpha} = 0$$

by (9.13) for $k = 1, \cdots, n$. Let $x_1 = 0$ and $x_2 = \dfrac{1}{2|n_2|}$. Then

$$f\left(0, \frac{1}{2|n_2|}, 0, \cdots, 0\right) = \frac{C}{(2\pi)^\alpha} N^{-\alpha} \geqslant \frac{C}{(2\pi)^\alpha} n^{-\alpha/2}.$$

Hence the theorem follows.

**Notes**

Theorem 9.3: Cf. N. M. Korobov [7].

Lemma 9.2: Cf. Wang Yuan [1].

Theorems of the type of Theorem 9.4 were first proved by V. S. Rjabenkii [1] and S. A. Smoljak [1] and their results were obtained by Theorem 9.7 with $O(p^{-\frac{\alpha}{2}+\frac{1}{4}+\varepsilon})$ and $N = [p^{1/2}]$ instead of $O(p^{\frac{-\alpha(2\alpha-1)}{4\alpha-1}+\varepsilon})$ and $N = [p^{\frac{2\alpha}{4\alpha-1}}]$ respectively. Theorem 9.7 was given by Wang Yuan [1,2].

Theorem 9.8: Cf. Hua Loo Keng and Wang Yuan [6,7].

Theorem 9.11: Cf. N. M. Korobov [6,7].

Theorems 9.14 and 9.15 were proved by Hua Loo Keng and Wang Yuan [7] and N. M. Korobov [7] respectively.

# Chapter 10

# Approximate Solution of Integral Equations and Differential Equations

## 10.1 Several lemmas

**Lemma 10.1** *If* $\sum_{i=1}^{s}\sum_{j=1}^{s}\alpha_{i_j}x_ix_j(\alpha_{i_j} = \alpha_{j^i})$ *is a semi-positive definite quadratic form, then*

$$0 \leqslant \det(\alpha_{i_j}) \leqslant \prod_{i=1}^{s}\alpha_{ii}.$$

*Proof.* Since the matrix $(\alpha_{ij})(1 \leqslant i, j \leqslant s)$ may be reduced to diagonal form

$$(\alpha_{i_j}) = \Lambda(\gamma_{i_j}\delta_{i_j})\Lambda', \quad 1 \leqslant i, j \leqslant s,$$

where $\gamma_{ii} \geqslant 0(1 \leqslant i \leqslant s)$ and

$$\Lambda = \begin{pmatrix} 1 & 0 & 0\cdots 0 \\ \lambda_{21} & 1 & 0\cdots 0 \\ & & \cdots \\ \lambda_{s1} & \lambda_{s2} & \lambda_{s3}\cdots 1 \end{pmatrix}$$

(Cf. Hua Loo Keng [2], Chap. 14), we have

$$\det(\alpha_{ij}) = \prod_{i=1}^{s}\gamma_{ii}$$

and

$$\alpha_{ii} = \sum_{j=1}^{s}\sum_{k=1}^{s}\lambda_{i_j}\gamma_{j^k}\delta_{j^k}\lambda_{ik} = \gamma_{ii} + \sum_{j<i}\lambda_{i_j}^2\gamma_{jj} \geqslant \gamma_{ii}$$

The lemma follows.

From Lemma 10.1, Hadamard's inequality can be easily derived.

**Lemma 10.2** *Let $\beta_{i_j}(1 \leqslant i,j \leqslant s)$ be real numbers. Then*

$$|\det(\beta_{i_j})| \leqslant \prod_{i=1}^{s}\left(\sum_{j=1}^{s}\beta_{i_j}^2\right)^{1/2}.$$

*Proof.* Let

$$\sum_{k=1}^{s}\left(\sum_{i=1}^{s}\beta_{ik}x_i\right)^2 = \sum_{i=1}^{s}\sum_{j=1}^{s}\left(\sum_{k=1}^{s}\beta_{ik}\beta_{j^k}\right)x_ix_j = \sum_{i=1}^{s}\sum_{j=1}^{s}\alpha_{i_j}x_ix_j.$$

Then

$$\det(\alpha_{i_j}) = (\det(\beta_{i_j}))^2, \quad 1 \leqslant i,j \leqslant s$$

and

$$\alpha_{ii} = \sum_{k=1}^{s}\beta_{i^k}^2, \quad 1 \leqslant i \leqslant s.$$

Hence the lemma follows by Lemma 10.1.

The geometrical meaning of Lemma 10.2 is that the volume of a parallelepiped does not exceed the product of the lengthes of its edges.

**Lemma 10.3** *Let $a_{i_j}(1 \leqslant i,j \leqslant s)$ be real numbers and $\tilde{A}_s(k) = \det(a'_{i_j})(1 \leqslant i,j \leqslant s)$, where $0 \leqslant k \leqslant s$ and*

$$a'_{i_j} = \begin{cases} 1 + a_{ii}, & if\ j = i, 1 \leqslant i \leqslant k, \\ a_{i_j}, & otherwise. \end{cases}$$

*Let $A_s(k)$ denote the upper bound of $|\tilde{A}_s(k)|$ under the condition $|a_{i_j}| \leqslant \gamma/s(1 \leqslant i,j \leqslant s)$, where $\gamma$ is a constant. Then there exist two constants $\gamma_1$ and $\gamma_2$ such that*

$$A_s(s-1) \leqslant \frac{\gamma_1}{s}, \quad A_s(s) \leqslant \gamma_2.$$

*Proof.* We express the determinant $\tilde{A}_s(k)$ as a sum of two determinants $A$ and $B$, where $(1,0,\cdots,0)$ and $(a_{11},a_{12},\cdots,a_{1s})$ are the first rows of $A$ and $B$ respectively and the other elements of $A$ and $B$ are the same as the corresponding elements of $\tilde{A}_s(k)$. Then

$$A_s(k) \leqslant A_{s-1}(k-1) + A_s(k-1).$$

If $k \geqslant 2$, then

$$A_{s-1}(k-1) \leqslant A_{s-2}(k-2) + A_{s-1}(k-2)$$

and

$$A_s(k-1) \leqslant A_{s-1}(k-2) + A_s(k-2).$$

Hence

$$A_s(k) \leqslant A_{s-2}(k-2) + 2A_{s-1}(k-2) + A_s(k-2).$$

Consequently

$$A_s(k) \leqslant A_{s-k}(0) + C_1^k A_{s-k+1}(0) + \cdots + A_s(0).$$

Take $k = s-1$. Then

$$A_s(s-1) \leqslant \sum_{v=1}^{s} C_{v-1}^{s-1} A_v(0).$$

Since $|a_{i_j}| \leqslant \gamma/s$, we have

$$A_v(0) \leqslant v^{v/2}\left(\frac{\gamma}{s}\right)^v$$

by Lemma 10.2 Hence

$$A_s(s-1) \leqslant \sum_{v=1}^{s} C_{v-1}^{s-1} v^{v/2}\left(\frac{\gamma}{s}\right)^v \leqslant \frac{1}{s}\sum_{v=1}^{\infty}\frac{\gamma^v v^{v/2}}{(v-1)!} = \frac{\gamma_1}{s}.$$

Using Lemma 10.2 again, we have

$$\begin{aligned} A_s(s) &\leqslant ((|1+a_{11}|^2 + \cdots + |a_{1s}|^2)\cdots(|a_{s1}|^2 + \cdots + |1+a_{ss}|^2))^{1/2} \\ &\leqslant \left(\left(1+\frac{\gamma}{s}\right)^2 + (s-1)\frac{\gamma^2}{s^2}\right)^{s/2} = \left(1+\frac{2\gamma+\gamma^2}{s}\right)^{s/2} \\ &\leqslant e^{r+\frac{\gamma^2}{2}} = \gamma_2. \end{aligned}$$

The lemma is proved.

For simplicity, we use capital Latin letters to denote the $s$-dimensional vector.

**Lemma 10.4** *Suppose that* $F(Q_1, \cdots, Q_r) \in H_{rs}^{\alpha}(C)(\alpha > 0)$. *If the quadrature formula given by the set* $M_k(1 \leqslant k \leqslant n)$

$$\int_{G_{rs}} F(Q_1, \cdots, Q_r) dQ_1 \cdots dQ_r = \frac{1}{n^r}\sum_{k_1,\cdots,k_r=1}^{n} F(M_{k_1}, \cdots, M_{k_r}) + O(\varepsilon(n))$$

*holds for* $r=1$, *where* $\varepsilon(n) = o(1)$(as $n \to \infty$) *and the constant implied by the symbol "O" depends on* $C, \alpha, r, s$ *only, then it holds for* $r > 1$ *also.*

*Proof.* We shall prove the lemma by induction. The lemma is true for $r = 1$ by the assumption. Suppose that $r \geqslant 2$ and the lemma is true for integers less than $r$. Since $F(Q_1, \cdots, Q_r) \in H^{\alpha}_{(r-1)_s}(C)$ for given $Q_r$ and $F(Q_1, \cdots, Q_r) \in H^{\alpha}_{s}(C)$ for given $Q_1, \cdots, Q_{r-1}$, we have

$$\begin{aligned}
&\int_{G_{rs}} F(Q_1, \cdots, Q_r) dQ_1 \cdots dQ_r \\
&= \int_{G_s} dQ_r \int_{G_{(r-1)s}} F(Q_1, \cdots, Q_{r-1}, Q_r) dQ_1 \cdots dQ_{r-1} \\
&= \frac{1}{n^{r-1}} \sum_{k_1, \cdots, k_{r-1}=1}^{n} \int_{G_s} F(M_{k_1}, \cdots, M_{k_{r-1}}, Q_r) dQ_r + O(\varepsilon(n)) \\
&= \frac{1}{n^r} \sum_{k_1, \cdots, k_r=1}^{n} F(M_{k_1}, \cdots, M_{k_r}) + O(\varepsilon(n)).
\end{aligned}$$

and the lemma follows by induction.

## 10.2 The approximate solution of the Fredholm integral equation of second type

In this section, we shall study the problem of approximate solution of the Fredholm integral equation of second type

$$\varphi(P) = \lambda \int_{G_s} K(P, Q) \varphi(Q) dQ + f(P), \tag{10.1}$$

where $f \in H^{\alpha}_{s}(C)$ and $K \in H^{\alpha}_{2s}(C)$.

Let

$$D(\lambda) = 1 + \sum_{v=1}^{\infty} \frac{(-1)^v}{v!} \lambda^v \int_{G_v s} K \begin{pmatrix} P_1, \cdots, P_v \\ P_1, \cdots, P_v \end{pmatrix} dP_1 \cdots dP_v$$

denote the Fredholm kernel of the equation (10.1), where

$$K \begin{pmatrix} P_1, \cdots, P_v \\ Q_1, \cdots, Q_v \end{pmatrix} = \det(K(P_i, Q_j)), \quad 1 \leqslant i, j \leqslant v.$$

Let

$$\Delta(\lambda) = \det\left(\delta_{ij} - \frac{\lambda}{n} K(M_i, M_j)\right), \quad 1 \leqslant i, j \leqslant n.$$

We suppose that

$$D(\lambda) \neq 0.$$

**Theorem 10.1** *Suppose that*

$$\sup_{f\in H_s^\alpha(C)}\left|\int_{G_s}F(P)dP-\frac{1}{n}\sum_{k=1}^{n}F(M_k)\right|\leqslant Cc(\alpha,s)\varepsilon(n),$$

*where* $\varepsilon(n)=o(1)(\text{as } n\to\infty)$. *Let* $\tilde{\varphi}(M_k)(1\leqslant k\leqslant n)$ *denote the solution of the system of linear equations*

$$\tilde{\varphi}(M_j)=\frac{\lambda}{n}\sum_{k=1}^{n}K(M_j,M_k)\tilde{\varphi}(M_k)+f(M_j),\quad 1\leqslant j\leqslant n. \tag{10.2}$$

*Then the solution of* (10.1) *may be expressed by*

$$\varphi(P)=f(P)+\frac{\lambda}{n}\sum_{k=1}^{n}K(P,M_k)\tilde{\varphi}(M_k)+O(\varepsilon(n)),$$

*where the constant implied by the symbol "O" depends on* $\lambda, K$ *and* $f$ *only.*

To prove Theorem 10.1, we shall need

**Lemma 10.5** *Let*

$$D_r(\lambda)=1+\sum_{v=1}^{r}\frac{(-1)^v}{v!}\lambda^v\int_{G_{vs}}K\begin{pmatrix}P_1,\cdots,P_v\\P_1,\cdots,P_v\end{pmatrix}dP_1\cdots dP_v.$$

*Then*

$$|D_r(\lambda)-D(\lambda)|\leqslant\frac{1}{2^r},$$

*if* $r$ *is sufficiently large.*

*Proof.* By Lemma 10.2,

$$\left|K\begin{pmatrix}P_1,\cdots,P_v\\P_1,\cdots,P_v\end{pmatrix}\right|\leqslant v^{v/2}C^v.$$

Take $r$ sufficiently large such that

$$r\geqslant(2e|\lambda|C)^2\quad\text{and}\quad\frac{3}{2^r}\leqslant\frac{1}{2}|D(\lambda)|. \tag{10.3}$$

Then

$$\begin{aligned}|D(\lambda)-D_r(\lambda)|&=\left|\sum_{v=r+1}^{\infty}\frac{(-1)^v}{v!}\lambda^v\int_{G_{vs}}K\begin{pmatrix}P_1,\cdots,P_v\\P_1,\cdots,P_v\end{pmatrix}dP_1\cdots dP_v\right|\\&\leqslant\sum_{v=r+1}^{\infty}\frac{(|\lambda|C)^v v^{v/2}}{v!}.\end{aligned}$$

Since $r! > r^r e^{-r}$ and

$$\frac{v!(|\lambda|C)^{v+1}(v+1)^{\frac{v+1}{2}}}{(v+1)!(|\lambda|C)^v v^{v/2}} = \frac{|\lambda|C}{\sqrt{v+1}}\left(1+\frac{1}{v}\right)^{v/2} \leqslant \frac{|\lambda|C\sqrt{e}}{\sqrt{r+2}} \leqslant \frac{1}{2\sqrt{e}} < \frac{1}{2}$$

for $v \geqslant r+1$, therefore

$$|D(\lambda) - D_r(\lambda)| \leqslant \frac{(|\lambda|C)^{r+1}(r+1)^{\frac{r+1}{2}}}{(r+1)!}\sum_{v=0}^{\infty}\frac{1}{2^v} \leqslant \frac{2(e|\lambda|C)^{r+1}}{(r+1)^{\frac{r+1}{2}}} \leqslant \frac{1}{2^r},$$

The lemma is proved.

**Lemma 10.6** *There exists constant $n_0 = n_0(\lambda, K, f)$ such that*

$$|\Delta(\lambda)| \geqslant \frac{1}{2}|D(\lambda)|$$

*for $n > n_0$.*

*Proof.* Choose $r$ satisfying (10.3). Let

$$\Delta_r(\lambda) = 1 + \sum_{v=1}^{r}\frac{(-1)^v}{v!}\lambda^v\frac{1}{n^v}\sum_{k_1,\cdots,k_v=1}^{n} K\left(\begin{array}{c} M_{k_1},\cdots,M_{k_v} \\ M_{k_1},\cdots,M_{k_v}\end{array}\right).$$

Then in a way similar to the proof of Lemma 10.5, we have

$$|\Delta(\lambda) - \Delta_r(\lambda)| \leqslant \left|\sum_{v=r+1}^{\infty}\frac{(-1)^v}{v!}\lambda^v\frac{1}{n^v}\sum_{k_1,\cdots,k_v=1}^{n} K\left(\begin{array}{c} M_{k_1}\cdots,M_{k_v} \\ (M_{k_1},\cdots,M_{k_v})\end{array}\right)\right| \leqslant \sum_{v=r+1}^{\infty}\frac{(|\lambda|C)^v \nu^{v/2}}{v!} \leqslant \frac{1}{2^r}.$$

The quadrature formula

$$\int_{G_{vs}} K\left(\begin{array}{c} P_1,\cdots,P_v \\ P_1,\cdots,P_v\end{array}\right)dP_1\cdots dP_v = \frac{1}{n^v}\sum_{k_1,\cdots,k_v=1}^{n} K\left(\begin{array}{c} M_{k_1},\cdots,M_{k_v} \\ M_{k_1},\cdots,M_{k_v}\end{array}\right) + O(\varepsilon(n))$$

holds for $v = 1$ by the assumption of the lemma. Since

$$K\left(\begin{array}{c} P_1,\cdots,P_v \\ P_1,\cdots,P_v\end{array}\right) \in H^{\alpha}_{vs}(Cc(v,\alpha,x)) \quad \text{for} \quad K(P,Q) \in H^{\alpha}_{2s}(C),$$

therefore it holds for $v > 1$ by Lemma 10.4. Hence

$$D_r(\lambda) - \Delta_r(\lambda) = \sum_{v=1}^{r} \frac{(-1)^v \lambda^v}{v!} \left( \int_{G_{vs}} K \begin{pmatrix} P_1, \cdots, P_v \\ P_1, \cdots, P_v \end{pmatrix} dP_1 \cdots dP_v \right.$$
$$\left. - \frac{1}{n^v} \sum_{k_1, \cdots, k_v = 1}^{n} K \begin{pmatrix} M_{k_1}, \cdots, M_{k_v} \\ M_{k_1}, \cdots, M_{k_v} \end{pmatrix} \right) = O(\varepsilon(n)),$$

where the constant implied by the symbol "$O$" depends on $\lambda, K, f$ only, i.e., there exists constant $n_0 = n_0(\lambda, K, f)$ such that

$$|D_r(\lambda) - \Delta_r(\lambda)| \leqslant \frac{1}{2^r}$$

for $n > n_0$. Then by Lemma 10.5,

$$|D(\lambda) - \Delta(\lambda)| \leqslant |D(\lambda) - D_r(\lambda)| + |D_r(\lambda) - \Delta_r(\lambda)|$$
$$+ |\Delta_r(\lambda) - \Delta(r)| \leqslant \frac{3}{2^r} \leqslant \frac{1}{2}|D(r)|.$$

Hence

$$|\Delta(\lambda)| \geqslant |D(\lambda)| - |D(\lambda)| - \Delta(\lambda)| \geqslant \frac{1}{2}|D(\lambda)|.$$

The lemma is proved.

**Lemma 10.7.** *Let $\varphi(P)$ denote the solution of* (10.1). *Then*

$$\varphi(P) \in H_s^\alpha(c(\lambda, K, f)).$$

*Proof.* Since

$$\varphi(P) - f(P) = \lambda \int_{G_s} K(P, Q)\varphi(Q)dQ,$$

therefore

$$|(\varphi(P) - f(P))^{(\alpha, \cdots, \alpha)}| \leqslant \sup_{Q \in G_s} |K(P, Q)^{(\alpha, \cdots, \alpha)}| \, |\lambda| \int_{G_s} |\varphi(Q)|dQ$$
$$\leqslant c(\lambda, K, f).$$

Hence

$$\varphi(P) - f(P) \in H_s^\alpha(c(\lambda, K, f))$$

and so

$$\varphi(P) = f(P) + (\varphi(P) - f(P)) \in H_s^\alpha(c(\lambda, K, f)).$$

The lemma is proved.

The proof of Theorem 10.1. It follows by Lemma 10.7 that

$$K(P,Q)\varphi(Q) \in H_s^\alpha(c(\lambda, K, f))$$

for given $P$. Hence

$$\varphi(P) = \frac{\lambda}{n}\sum_{k=1}^{n} K(P, M_k)\varphi(M_k) + f(P) + O(\varepsilon(n)) \tag{10.4}$$

and

$$\varphi(M_j) = \frac{\lambda}{n}\sum_{k=1}^{n} K(M_j, M_k)\varphi(M_k) + f(M_j) + O(\varepsilon(n)), \quad 1 \leqslant j \leqslant n.$$

From (10.2), we have the system of linear equations

$$z_j = \sum_{k=1}^{n} a_{jk} z_k + b_j, \quad 1 \leqslant j \leqslant n,$$

where

$$z_j = \varphi(M_j) - \tilde{\varphi}(M_j),$$

$$a_{jk} = \frac{\lambda}{n} K(M_j, M_k),$$

and

$$b_j = O(\varepsilon(n)).$$

Let $\Delta_k(\lambda)$ denote the determinant obtained from $\Delta(\lambda)$ by replacing its $k$-th column

$$\left(-\frac{\lambda}{n}K(M_1, M_k), \cdots, 1 - \frac{\lambda}{n}K(M_k, M_k), \cdots, -\frac{\lambda}{n}K(M_n, M_k)\right)'$$

by

$$(b_1, \cdots, b_n)'.$$

Then we have

$$z_j = \frac{\Delta_j(\lambda)}{\Delta(\lambda)}, \quad 1 \leqslant j \leqslant n.$$

When $n$ is sufficiently large, we have

$$|\Delta(\lambda)| > \frac{1}{2}|D(\lambda)| > 0$$

by Lemma 10.6. Further more since

$$\left|\frac{\lambda}{n}K(M_j, M_k)\right| \leqslant \frac{|\lambda| C}{n},$$

therefore by Lemma 10.3, we have

$$|\Delta_j(\lambda)| \leqslant |b_j B_j| + \sum_{\substack{1\leqslant k\leqslant n \\ k\neq j}} |b_k B_k|$$
$$\leqslant \gamma_2 |b_j| + \frac{\gamma_1}{n}\sum_{k=1}^{n} |b_k| = O(\varepsilon(n)),$$

where $B_k$ denotes the cofactor of $b_k$ in $\Delta_k(\lambda)$. Hence

$$z_j = O(\varepsilon(n)), \quad 1 \leqslant j \leqslant n.$$

Substituting into (10.4), the theorem follows.

Especially, let $M_k (1 \leqslant k \leqslant n)$ be the sets introduced in Chap. 4. We obtain various approximate formulas for the solutions of the equation (10.1).

## 10.3 The approximate solution of the Volterra integral equation of second type

In this section, we shall study the problem of approximate solution of the Volterra equation of second type

$$\varphi(x) = \int_0^x K(x,y)\varphi(y)dy + f(x), \tag{10.5}$$

where $f \in H_1^\alpha(C)$ and $K(x,y) \in H_2^\alpha(C)$.

Introduce the notations

$$\mu(\alpha) = \begin{cases} \dfrac{\alpha}{2}, & \text{if } \alpha > 1, \\ \dfrac{2\alpha^2}{1+4\alpha-\alpha^2}, & \text{if } 1 \geqslant \alpha > 0, \end{cases}$$
$$q = \left[p^{\frac{\mu(\alpha)}{\alpha}}\right],$$
$$Q = \left[\mu(\alpha)\frac{\log_2 p}{\log_2 \log_2 3p}\right]$$

and

$$B_{k,v,p}(a) = \sum_{\bar{m}_1\cdots\bar{m}_v<p} e^{-2\pi i(m_1+m_2a+\cdots+m_v a^{v-1})k/p}$$
$$\cdot \int_0^x \int_0^{x_1} \cdots \int_0^{x_{v-1}} e^{2\pi i(m_1x_1+\cdots+m_vx_v)}dx_1\cdots dx_v.$$

**Theorem 10.2** *There exists an integer $a(= a(p))$ such that the solution of* (10.5) *may be represented as*

$$\varphi(x) = f(x) + \frac{1}{p}\sum_{k=1}^{p}\sum_{v=1}^{Q} B_{k,v,p}(a) K\left(x, \frac{k}{p}\right) K\left(\frac{k}{p}, \frac{ak}{p}\right)\cdots$$
$$\cdots K\left(\frac{a^{v-2}k}{p}\frac{a^{v-1}k}{p}\right) f\left(\frac{a^{v-1}k}{p}\right) + O(p^{-\mu(\alpha)+\varepsilon}),$$

*where the constant implied by the symbol "O" depends only on $K, f, \varepsilon$.*

*Proof.* The solution of the equation (10.5) is given by the Neumann series

$$\varphi(x) = f(x) + \sum_{\nu=1}^{\infty} \varphi_v(x),$$

where

$$\varphi_v(x) = \int_0^x \int_0^{x_1} \cdots \int_0^{x_{v-1}} R_v dx_1 \cdots dx_v$$

and

$$R_v = R_v(x, x_1, \cdots, x_v) = K(x, x_1)K(x_1, x_2)\cdots K(x_{v-1}, x_v)f(x_v).$$

Since $R_v \in H^{\alpha}_{v+1}(2^{(\alpha+1)(v+1)}C^{v+1})$, we have

$$|\varphi_v(x)| \leqslant 2^{(\alpha+1)(v+1)}C^{v+1}\int_0^x \int_0^{x_1} \cdots \int_0^{x_{v-1}} dx_1 \cdots dx_v$$
$$\leqslant \frac{2^{(\alpha+1)(\alpha+1)}C^{v+1}}{v!}$$

and

$$\left|\sum_{v=Q+1}^{\infty} \varphi_v(x)\right| \leqslant \sum_{v=Q+1}^{\infty} \frac{2^{(\alpha+1)(v+1)}C^{v+1}}{v!} \leqslant \frac{c(C,\alpha)^Q}{Q!} \leqslant c(K, f, \varepsilon)p^{-\mu(\alpha)+\varepsilon}.$$

Let

$$S_v = S_v(x, x_1, \cdots, x_v)$$
$$= \frac{1}{p}\sum_{k=1}^{p} K\left(x, \frac{k}{p}\right) K\left(\frac{k}{p}, \frac{ak}{p}\right) \cdots K\left(\frac{a^{v-2}k}{p}, \frac{a^{v-1}k}{p}\right) f\left(\frac{a^{v-1}k}{p}\right)$$
$$\cdot \sum_{\bar{m}_1 \cdots \bar{m}_v < q} e^{-2\pi i(m_1+m_2a+\cdots+m_v a^{v-1})k/p + 2\pi i(m_1x_1+\cdots+m_v x_v)}.$$

Then it follow from Theorem 9.10 that there exists an integer $a(= a(p))$ such that

$$\|R_v - S_v\|_2 \leqslant C^{v+1}c(\alpha, \varepsilon)^{v+1}v!^{1/2}p^{-\mu(\alpha)+\varepsilon/2}$$

and

$$\begin{aligned}
&\left|\varphi_v(x) - \int_0^x \cdots \int_0^{x_{v-1}} S_v dx_1 \cdots dx_v\right| \\
\leqslant &\int_0^x \cdots \int_0^{x_{v-1}} |R_v - S_v| dx_1 \cdots dx_v \\
\leqslant &\left(\int_0^x \cdots \int_0^{x_{v-1}} dx_1 \cdots dx_v\right)^{1/2} \|R_v - S_v\|_2 \\
\leqslant &C^{v+1} c(\alpha, \varepsilon)^v p^{-\mu(\alpha)+\varepsilon/2}.
\end{aligned}$$

The theorem follows.

## 10.4 The eigenvalue and eigenfunction of the Fredholm equation

For $f(x) = 0$, the Fredholm equation of second type

$$\varphi(x) = \lambda \int_0^1 K(x, y)\varphi(y)dy \tag{10.6}$$

is called the homogeneous equation. If there is a $\lambda$ such that (10.6) has a non-zero solution $\varphi(x)$, then $\lambda$ is called the eigenvalue of the kernel $K(x, y)$ and $\varphi(x)$ the eigenfunction of $K(x, y)$ corresponding to $\lambda$. The maximum number of linear independent eigenfunctions over the complex number field corresponding to an eigenvalue $\lambda$ is called the multiplicity of $\lambda$.

In this section, we suppose that $K(x, y) > 0$ and $K(x, y) = K(y, x)$ for $(x, y) \in G_2$ and that $K(x, y) \in H_2^\alpha(C)$. We shall study the problem of the approximate solution of the least eigenvalue and its corresponding eigenfunction of the equation (10.6). First, We shall mention some well known results for the integral equation (Cf. V. S. Vladimirov [1]).

**Lemma 10.8** *The equation* (10.6) *has eigenvalues. The number of its eigenvalues is denumerable. The eigenvalues are all real and have no finite limit point. Moreover the multiplicity of every eigenvalue is finite.*

Now we arrange the eigenvalues of the equation (10.6) according to their absolute values

$$|\lambda_1| \leqslant |\lambda_2| \leqslant \cdots, \tag{10.7}$$

where if $\lambda$ has the multiplicity $k$, then it will appear $k$ times in (10.7). The corresponding eigenfunctions are denoted by

$$\varphi_1, \varphi_2, \cdots .$$

Without loss of generality, we may suppose that

$$\|\varphi_i\|_2 = 1, \quad i = 1, 2, \cdots .$$

**Lemma 10.9** *$\lambda_1$ is positive and simple and $\varphi_1(x) > 0 (x \in G_1)$.*

By Lemmas 10.8 and 10.9, we have

$$0 < \lambda_1 < |\lambda_2| \leqslant \cdots .$$

Suppose that $f \in H_1^\alpha(C)$ and $f$ is a non-negative real function satisfying $\|f\|_2 = 1$, for example $f(x) = 1$. Then it follows by Schwarz's inequality that

$$0 < c_1 = \int_0^1 f(x)\varphi_1(x)dx \leqslant 1.$$

Denote

$$\Phi_s(x) = \frac{\int_{G_s} R_s d\mathbf{x}}{\|R_s\|_2}, \quad s \geqslant 1$$

and

$$\Lambda_s = \frac{\|R_{s-1}\|_2}{\|R_s\|_2}, \quad s \geqslant 2,$$

where

$$\begin{aligned} R_s &= R_s(x, x_1, \cdots, x_s) \\ &= K(x, x_1)K(x_1, x_2) \cdots K(x_{s-1}, x_s) f(x_s). \end{aligned}$$

**Lemma 10.10**

$$0 \leqslant \Lambda_s - \lambda_1 \leqslant \frac{1 - c_1^2}{c_1^2} \cdot \frac{\lambda_1}{2} \left(\frac{\lambda_1}{|\lambda_2|}\right)^{2s-2}, \quad s \geqslant 2$$

and

$$\|\Phi_s - \varphi_1\|_2 \leqslant \frac{\sqrt{1 - c_1^2}}{c_1} \left(\frac{\lambda_1}{|\lambda_2|}\right)^s, \quad s \geqslant 1.$$

Introduce the notations

$$R_s^* = R_s^*(x) = \frac{1}{n}\sum_{k=1}^{n} R_s\left(x, \frac{a_1 k}{n}, \cdots, \frac{a_s k}{n}\right)$$

and

$$\tilde{R}_s = \left(\frac{1}{n}\sum_{k=1}^{n} R_s\left(\frac{a_1 k}{n}, \cdots, \frac{a_{s+1} k}{n}\right)^2\right)^{1/2}.$$

**Theorem 10.3** *Suppose that the congruence*

$$\sum_{i=1}^{s+1} a_i m_i \equiv 0 (\bmod n)$$

*has no solution in the domain*

$$\bar{m}_1 \cdots \bar{m}_{s+1} \leqslant M, \quad (m_1, \cdots, m_{s+1}) \neq (0, \cdots, 0).$$

*Then*

$$\left|\frac{\tilde{R}_{s-1}}{\tilde{R}_s} - \lambda_1\right| \leqslant \frac{1-c_1^2}{c_1^2}\cdot\frac{\lambda_1}{2}\left(\frac{\lambda_1}{|\lambda_2|}\right)^{2s-2} + c(K,f,\varepsilon)M^{-\alpha+\varepsilon}, \quad s \geqslant 2 \tag{10.8}$$

*and*

$$\left\|\frac{R_s^*(x)}{\tilde{R}_s} - \varphi_1(x)\right\|_2 \leqslant \frac{\sqrt{1-c_1^2}}{c_1}\left(\frac{\lambda_1}{|\lambda_2|}\right)^{s} + c(K,f,\varepsilon)M^{-\alpha+\varepsilon}, \quad s \geqslant 1. \tag{10.9}$$

*Proof.* Since

$$R_s \in H_{s+1}^{\alpha}(2^{(\alpha+1)(s+1)}C^{s+1}), \quad R_s^2 \in H_{s+1}^{\alpha}(2^{2(\alpha+1)(s+1)}C^{2(s+1)}),$$

we have

$$|\,\|R_s\|_2^2 - \tilde{R}_s^2| \leqslant c(K,f,\varepsilon)M^{-\alpha+\varepsilon}$$

by Theorem 7.9 and so

$$|\,\|R_s\|_2 - \tilde{R}_s| \leqslant c(K,f,\varepsilon)M^{-\alpha+\varepsilon}.$$

Since $\|R_s\|_2 = c(K,f) > 0$, we may suppose that $\tilde{R}_s = c(K,f) > 0$. Hence

$$\begin{aligned}\left|\frac{\tilde{R}_{s-1}}{\tilde{R}_s} - \Lambda_s\right| &= \frac{|\tilde{R}_s\|R_{s-1}\|_2 - \tilde{R}_{s-1}\|R_s\|_2|}{\tilde{R}_s\|R_s\|_2}\\ &\leqslant \frac{|\|R_s\|_2(\|R_{s-1}\|_2 - \tilde{R}_{s-1}) + \|R_{s-1}\|_2(\tilde{R}_s - \|R_s\|_2)|}{c(K,f)}\\ &\leqslant c(K,f,\varepsilon)M^{-\alpha+\varepsilon}.\end{aligned}$$

Consequently, (10.8) follows by Lemma 10.10 and

$$\left|\frac{\tilde{R}_{s-1}}{\tilde{R}_s} - \lambda_1\right| \leqslant \left|\frac{\tilde{R}_{s-1}}{\tilde{R}_s} - \Lambda_s\right| + |\Lambda_s - \lambda_1|.$$

By Theorem 7.9

$$\left\|\int_{G_s} R_s d\mathbf{x} - R_s^*\right\|_2 \leqslant \sup_{x\in G_1}\left|\int_{G_s} R_s d\mathbf{x} - R_s^*\right| \leqslant c(K, f, \varepsilon)M^{-\alpha+\varepsilon}$$

and so by Minkowski's inequality, we have

$$\begin{aligned}\left\|\frac{R_s^*}{\tilde{R}_s} - \Phi_s\right\|_2 &\leqslant c(K, f)\left\|\tilde{R}_s\int_{G_s} R_s d\mathbf{x} - \|R_s\|_2 R_s^*\right\|_2\\ &= c(K, f)\left\|\tilde{R}_s\left(\int_{G_s} R_s d\mathbf{x} - R_s^*\right) + R_s^*(\tilde{R}_s - \|R_s\|_2)\right\|_2\\ &\leqslant c(K, f)\left(\left\|\tilde{R}_s\left(\int_{G_s} R_s d\mathbf{x} - R_s^*\right)\right\|_2 + \left\|R_s^*(\tilde{R}_s - \|R_s\|_2)\right\|_2\right)\\ &\leqslant c(K, f)M^{-\alpha+\varepsilon}.\end{aligned}$$

Hence (10.9) follows by Lemma 10.10 and

$$\left\|\frac{R_s^*}{\tilde{R}_s} - \varphi_1\right\|_2 \leqslant \left\|\frac{R_s^*}{\tilde{R}_s} - \Phi_s\right\| + \|\Phi_s - \varphi_1\|_2.$$

The theorem is proved.

## 10.5 The Cauchy problem of the partial differential equation of the parabolic type

In this section, we shall study the approximate solution of the parabolic equation

$$\frac{\partial u}{\partial t} = \left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_s^2}\right)u,$$
$$0 \leqslant t \leqslant T, \quad -\infty < x_v < \infty (1 \leqslant v \leqslant s).$$

Suppose that the initial condition is

$$u(0, \mathbf{x}) = f(\mathbf{x}) \in E_s^\alpha(C).$$

where $\alpha > 1$.

**Theorem 10.4** *If the congruence*

$$(\mathbf{a}, \mathbf{m}) = \sum_{i=1}^{s} a_i m_i \equiv 0 \pmod{n} \tag{10.10}$$

*has no solution in the domain*

$$\|\mathbf{m}\| \leqslant M, \quad \mathbf{m} \neq \mathbf{0}, \tag{10.11}$$

*then*

$$\sup_{\mathbf{x}\in G_s}\left|u(t,\mathbf{x}) - \sum_{\|\mathbf{m}\|<N}\left(\frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right)e^{-\frac{2\pi i(\mathbf{a},\mathbf{m})k}{n}}\right)e^{-4\pi^2(\mathbf{m},\mathbf{m})t+2\pi i(\mathbf{m},\mathbf{x})}\right|$$
$$\leqslant Cc(\alpha,s,\varepsilon)M^{-\frac{\alpha(\alpha-1)}{2\alpha-1}+\varepsilon}.$$

*where* $N = \left[M^{\frac{\alpha}{2\alpha-1}}\right]$.

*Proof.* Let

$$u(t,\mathbf{x}) = \sum C(t,\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})}.$$

Then

$$\sum \frac{d}{dt}C(t,\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})} = \frac{\partial u}{\partial t} = \left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_s^2}\right)u$$
$$= -\sum C(t,\mathbf{m})4\pi^2(\mathbf{m},\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})}$$

and so by the comparison of the coefficients of $e^{2\pi i(\mathbf{m},\mathbf{x})}$, we have

$$\frac{d}{dt}C(t,\mathbf{m}) = -4\pi^2(\mathbf{m},\mathbf{m})C(t,\mathbf{m}),$$
$$\int \frac{dC(t,\mathbf{m})}{C(t,\mathbf{m})} = -4\pi^2(\mathbf{m},\mathbf{m})\int dt,$$
$$C(t,\mathbf{m}) = ce^{-4\pi^2(\mathbf{m},\mathbf{m})t}.$$

Since

$$C(0,\mathbf{m}) = C(\mathbf{m}),$$

where $C(\mathbf{m})$ is the Fourier coefficient of $f(\mathbf{x})$, therefore $c = C(\mathbf{m})$ and

$$C(t,\mathbf{m}) = C(\mathbf{m})e^{-4\pi^2(\mathbf{m},\mathbf{m})t}.$$

Hence

$$u(t,\mathbf{x}) = \sum C(\mathbf{m})e^{-4\pi^2(\mathbf{m},\mathbf{m})t+2\pi i(\mathbf{m},\mathbf{x})}$$

and

$$\sup_{\mathbf{x}\in G_s}\left|u(t,\mathbf{x}) - \sum_{\|\mathbf{m}\|<N}\left(\frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right)e^{-\frac{2\pi i(\mathbf{a},\mathbf{m})k}{n}}\right)e^{-4\pi(\mathbf{m},\mathbf{m})t+2\pi i(\mathbf{m},\mathbf{x})}\right|$$
$$\leqslant \sum\nolimits_1 + \sum\nolimits_2,$$

where

$$\sum\nolimits_1 = \sup_{f\in E_s(C)} \sum_{\|\mathbf{m}\|<N} \left| C(\mathbf{m}) - \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) e^{-\frac{2\pi i(\mathbf{a},\mathbf{m})k}{n}} \right|$$

and

$$\sum\nolimits_2 = \sup_{f\in E_s^\alpha(C)} \sum_{\|\mathbf{m}\|\geqslant N} |C(\mathbf{m})|.$$

By the argument of the proof of Theorem 9.5, the theorem follows.

## 10.6 The Dirichlet problem of the partial differential equation of the elliptic type

In this section, we shall study the approximate solution of the elliptic equation

$$\left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_s^2}\right) u = f. \tag{10.12}$$

Suppose that $f \in E_s^\alpha(C)(\alpha > 1)$ and $f$ is an odd function with respect to each variable and that $u(\mathbf{x}) = 0$ if $\mathbf{x}$ belongs to the boundary of $G_s$.

Introduce the notations

$$0 \leqslant \omega \leqslant 1,$$
$$v(\alpha,\omega) = \frac{\alpha(\alpha+\omega-1)}{2\alpha-1}$$

and

$$g(\mathbf{x}) = \sum B(\mathbf{m})C(\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})}, \tag{10.13}$$

where $C(\mathbf{m})$ is the Fourier coefficient of $f$ and $B(\mathbf{m})$ satisfies

$$|B(\mathbf{m})| \leqslant \frac{1}{\|\mathbf{m}\|^\omega}.$$

**Theorem 10.5** *Suppose that $s \geqslant 2$. If the congruence* (10.10) *had no solution in* (10.11), *then*

$$\sup_{\mathbf{x}\in G_s} \left| u(\mathbf{x}) - \sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right)\psi_k(\mathbf{x}) \right| \leqslant Cc(\alpha,s,\varepsilon)M^{v(\alpha,\frac{2}{s})+\varepsilon},$$

*where*

$$\psi_k(\mathbf{x}) = -\frac{1}{4\pi^2 n} \sum_{\|\mathbf{m}\|<N}{}' \frac{e^{2\pi i\left(\mathbf{m},\mathbf{x}-\frac{k\mathbf{a}}{n}\right)}}{(\mathbf{m},\mathbf{m})}, \quad 1 \leqslant k \leqslant n$$

*and* $N = \left[M^{\frac{\alpha}{2\alpha-1}}\right]$.

To prove the theorem, we shall need

**Lemma 10.11** *If* $a_i \geqslant 0 (1 \leqslant i \leqslant s)$, *then*

$$(a_1 \cdots a_s)^{1/s} \leqslant \frac{a_1 + \cdots + a_s}{s}$$

(Cf. Hua Loo Keng [2], Chap. 20).

**Lemma 10.12** *If the congruence* (10.10) *has no solution in* (10.11), *then*

$$\begin{aligned}\mathfrak{S} = \sup_{\mathbf{x}\in G_s} &\left| g(\mathbf{x}) - \sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{kn}\right) \chi_k(\mathbf{x}) \right| \\ &\leqslant Cc(\alpha,\omega,s,\varepsilon) M^{-v(\alpha,\omega)+\varepsilon},\end{aligned}$$

*where*

$$\chi_k(\mathbf{x}) = \frac{1}{n} \sum_{\|\mathbf{m}\|<N} B(\mathbf{m}) e^{2\pi i(\mathbf{m},\mathbf{x}-\frac{k\mathbf{a}}{n})}, \quad 1 \leqslant k \leqslant n.$$

*Proof* By (10.13), we have

$$\mathfrak{S} \leqslant \sum\nolimits_1 + \sum\nolimits_2,$$

where

$$\sum\nolimits_1 = \sum_{\|\mathbf{m}\|<N} |B(\mathbf{m})| \left| C(\mathbf{m}) - \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) e^{\frac{-2\pi i(\mathbf{a},\mathbf{m})k}{n}} \right|$$

and

$$\sum\nolimits_2 = \sum_{\|\mathbf{m}\|\geqslant N} |B(\mathbf{m})||C(\mathbf{m})|.$$

Since

$$\begin{aligned}&\left| C(\mathbf{m}) - \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k\mathbf{a}}{n}\right) e^{\frac{-2\pi i(\mathbf{a},\mathbf{m})k}{n}} \right| \\ &\leqslant C \sum_{(\mathbf{a},\mathbf{l})\equiv 0 \pmod n}{}' \frac{1}{\|\mathbf{l}+\mathbf{m}\|^{\alpha}}\end{aligned}$$

(Cf. §9.4), we have

$$\begin{aligned}\sum\nolimits_1 &\leqslant C\sum_{k=0}^{[\log_2 N]}\sum_{2^{-k-1}\leqslant\|\mathbf{m}\|<2^{-k}N}\frac{1}{\|\mathbf{m}\|^{\omega}}\sum_{(\mathbf{a},\mathbf{l})\equiv 0 \pmod n}{}' \frac{1}{\|\mathbf{l}+\mathbf{m}\|^{\alpha}}\\
&\leqslant C\sum_{k=0}^{[\log_2 N]}(2^{-k-1}N)^{-\omega}\sum_{\|\mathbf{m}\|<2^{-k}N}\sum_{(\mathbf{a},\mathbf{l})\equiv 0 \pmod n}{}' \frac{1}{\|\mathbf{l}+\mathbf{m}\|^{\alpha}}\\
&\leqslant Cc(\alpha,s)N^{\alpha-\omega}\sum_{k=0}^{[\log_2 N]}2^{-(\alpha-\omega)k}\sum_{(\mathbf{a},\mathbf{l})\equiv 0 \pmod n}{}' \frac{1}{\|\mathbf{l}\|^{\alpha}}\\
&\leqslant Cc(\alpha,s,\varepsilon)N^{\alpha-\omega}M^{-\alpha+\varepsilon}\sum_{k=0}^{\infty}2^{-(\alpha-\omega)k}\\
&\leqslant Cc(\alpha,\omega,s,\varepsilon)M^{-v(\alpha,\omega)+\varepsilon}\end{aligned}$$

by Theorem 7.9 and Lemma 9.2. Take $\varepsilon<\alpha-1$. Then

$$\begin{aligned}\sum\nolimits_2 &\leqslant C\sum_{\|\mathbf{m}\|\geqslant N}\frac{1}{\|\mathbf{m}\|^{\alpha+\omega}}\leqslant C\sum_{\|\mathbf{m}\|\geqslant N}\frac{1}{\|\mathbf{m}\|^{\alpha+\omega-1-\varepsilon}\|\mathbf{m}\|^{1+\varepsilon}}\\
&\leqslant CN^{-\alpha-\omega+1+\varepsilon}\sum\frac{1}{\|\mathbf{m}\|^{1+\varepsilon}}\leqslant Cc(s,\varepsilon)M^{-v(\alpha,\omega)+\varepsilon}.\end{aligned}$$

The lemma follows.

The proof is of Theorem 10.5. Let

$$g(\mathbf{x})=\sum B(\mathbf{m})C(\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})},$$

where $C(\mathbf{m})$ is the Fourier coefficient of $f$ and

$$B(\mathbf{m})=\begin{cases}0, & \text{if } \mathbf{m}=\mathbf{0},\\ -\dfrac{1}{4\pi^2(\mathbf{m},\mathbf{m})}, & \text{if } \mathbf{m}\neq\mathbf{0}.\end{cases}\tag{10.14}$$

We shall prove that $g(\mathbf{x})$ is the solution of (10.12). Since $f(\mathbf{x})$ is an odd function with respect to each variable, therefore

$$C(\mathbf{0})=0$$

and

$$C(m_1,\cdots,m_v,\cdots,m_s)=-C(m_1,\cdots,-m_v,\cdots,m_s).$$

It follows that

$$\left(\frac{\partial^2}{\partial x_1^2}+\cdots+\frac{\partial^2}{\partial x_s^2}\right)g=\sum{}' C(\mathbf{m})e^{2\pi i(\mathbf{m},\mathbf{x})}=f(\mathbf{x})$$

and

$$\begin{aligned} g(x_1,\cdots,-x_v,\cdots,x_s) &= -\frac{1}{4\pi^2}\sum{}' \frac{C(\mathbf{m})}{(\mathbf{m},\mathbf{m})}e^{2\pi i(\mathbf{m},\mathbf{x})-4\pi i m_v x_v} \\ &= \frac{1}{4\pi^2}\sum{}' \frac{C(\mathbf{m})}{(\mathbf{m},\mathbf{m})}e^{2\pi i(\mathbf{m},\mathbf{x})}=-g(\mathbf{x}). \end{aligned}$$

Hence $g(\mathbf{x})$ is a solution of (10.12) and vanishes on the boundary of $G_s$, i.e.,

$$u(\mathbf{x})=g(\mathbf{x}).$$

For $\mathbf{m}\neq \mathbf{0}$, we have

$$(\mathbf{m},\mathbf{m})\geqslant s\|\mathbf{m}\|^{2/s}$$

by Lemma 10.11 and so

$$|B(\mathbf{m})|\leqslant \frac{1}{4\pi^2 s\|\mathbf{m}\|^{2/s}}.$$

by (10.14). Hence the theorem follows by Lemma 10.12.

## 10.7 Several remarks

1. We may also use the method of §10.2 to treat the problem of the approximate solution of the Fredholm integral equation of second type, if $K\in B_{2s}$ and $f\in B_s$.

2. The method of §10.3 may be used to treat the problem of the approximate solution of integral equation of the type

$$\begin{aligned} \varphi(x_1,\cdots,x_{s+l}) = \int_0^1\cdots\int_0^1\int_0^{x_{s+1}}\cdots\int_0^{x_{s+l}} K(x_1,\cdots,x_{s+l},y_1,\cdots,y_{s+l}) \\ \cdot\varphi(y_1,\cdots,y_{s+l})dy_1\cdots dy_{s+l}+f(x_1,\cdots,x_{s+l}), \end{aligned}$$

where $s\geqslant 0$ and $l\geqslant 1$ and where $f\in H^\alpha_{s+l}(C)$ and $K\in H^\alpha_{2(s+l)}(C)$ (or $f\in B_{s+l}$ and $K\in B_{2(s+l)}$).

3. We may also use the method of §10.3 to treat the problem of the approximate solution of the Fredholm equation of second type, if $\lambda$ is sufficiently small.

4. The method of §10.3 may be used also to treat the problem of the approximate solution of the linear parabolic equation

$$\frac{\partial u(t,\mathbf{x})}{\partial x} = \left(\sum_{i=1}^{s} \frac{\partial^2}{\partial x_i^2}\right) u(t,\mathbf{x}) + \sum_{i=1}^{s} \boldsymbol{a}_i(t,\mathbf{x}) \frac{\partial u(t,\mathbf{x})}{\partial x_i}$$
$$+\boldsymbol{a}(t,\mathbf{x})u(t,\mathbf{x}) + f(t,\mathbf{x}),$$
$$0 \leqslant t \leqslant T, \quad -\infty < x_i < \infty (1 \leqslant i \leqslant s),$$

where $u(0,\mathbf{x}) = 0$ and $\boldsymbol{a}_i$, $\boldsymbol{a}$ and $f$ belong to $H_{s+1}^{\alpha}(C)$.

Since the solution of the partial differential equation may be represented as the solution of the Volterra integral equation, it is given by the well-known Neumann series and so it may be represented approximately by a fintite sum.

5. The method of §10.4 may be used to treat the problem of the approximate solution of the eigenvalue and eigenfunction of the homogeneous Fredholm integral equation of second type

$$\varphi(P) = \lambda \int_{G_s} K(P,Q)\varphi(Q)dQ,$$

where $K(P,Q) \in H_{2s}^{\alpha}(C)$ or $K(P,Q) \in B_{2s}$.

6. The number theoretic method may be used also to arrange the experimental design and to find the optimal points of a function in a bounded reqion.

## Notes

Theorem 10.1 was first proved by N. M. Korobov [3, 7] for the functions $f \in E_s^{\alpha}(C)$ and $K \in E_{2s}^{\alpha}(C)$, where $\alpha > 1$ (Cf. also I. F. Sarygin [1], Wang Yuan [2] and Hua Loo Keng and Wang Yuan [3, 6, 7]).

Theorem 10.2 is an improvement of an earlier theorem of Yu. N. Sahov [1, 3] (Cf. Wang Yuan [1, 2] and Hua Loo Keng and Wang Yuan [3, 6, 7]).

Concerning the eigewalues and eigenfunctions of the Fredholm equation, we refer also Yu. N. Sahov [2].

Theorem 10.5 is an improvement of an earlier theorem of N. M. Korobov [7] (Cf. Wang Yuan [3, 4]).

Concerning the problems stated in §10.7, we refer N. M. Korobov [7], Hua Loo Keng and Wang Yuan [3, 6, 7], E. Hlawka [4, 5], V. S. Rjabenkii [2], Wang Yuan, Zhu Yao Cheng and Jian Yun Cui [1], V. T. Stojancev [1], Xu Guang Shan [1] and Wang Yuan and Fang Kai Tai [1].

# Appendix Tables

We use the notations $W_l(n, \mathbf{h})$ (Cf. §8.1) and $\rho(n, \mathbf{h}) = \min_m \|\mathbf{m}\|$, where $\mathbf{m}$ runs over the integral vectors with $\mathbf{m} \neq \mathbf{0}$ and $(\mathbf{h}, \mathbf{m}) \not\equiv 0 \pmod{n}$.

Table (1) is given by the Fibonacci sequence $F_m (= F_{2,m})$, i.e.

$$n = F_m, \quad \mathbf{h} = (1, F_{m-1}).$$

Tables (2)—(12) are given by the methods of Chap. 8 (Cf. §8.8). In table (11), the $\mathbf{h}(n)'s$ coresponding to $s = 12$ and 13 are obtained by neglecting the components $h_{13}, h_{14}$ and $h_{14}$ respectively and we use the notation $W_2(s, n, \mathbf{h})$ instead of $W_2(n, \mathbf{h})$.

**1** $(s = 2, h_1 = 1, h_2 = F_{m-1}, n = F_m)^*$

| $n$ | 13 | 21 | 34 | 55 | 89 |
|---|---|---|---|---|---|
| $W_2(n, \mathbf{h})$ | $4.7586 \times 10^{-1}$ | $2.0909 \times 10^{-1}$ | $8.9745 \times 10^{-2}$ | $3.8148 \times 10^{-2}$ | $1.6033 \times 10^{-2}$ |
| $n$ | 144 | 233 | 377 | 610 | 987 |
| $W_2(n, \mathbf{h})$ | $6.6851 \times 10^{-3}$ | $2.7673 \times 10^{-3}$ | $1.1388 \times 10^{-3}$ | $4.6619 \times 10^{-4}$ | $1.8900 \times 10^{-4}$ |
| $n$ | 1, 597 | 2, 584 | 4, 181 | 6, 765 | 10, 946 |
| $W_2(n, \mathbf{h})$ | $7.7127 \times 10^{-5}$ | $3.1200 \times 10^{-5}$ | $1.2581 \times 10^{-5}$ | $5.0595 \times 10^{-6}$ | $2.0293 \times 10^{-6}$ |
| $n$ | 17, 711 | 28, 657 | 46, 368 | 75, 025 | |
| $W_2(n, \mathbf{h})$ | $8.1206 \times 10^{-7}$ | $3.2376 \times 10^{-7}$ | $1.2819 \times 10^{-7}$ | $5.1270 \times 10^{-8}$ | |

**2** $(s = 3, h_1 = 1)$

| $n$ | $h_2$ | $h_3$ | $\rho(n, \mathbf{h})$ | $W_2(n, \mathbf{h})$ | $W_4(n, \mathbf{h})$ |
|---|---|---|---|---|---|
| 21 | 3 | 8 | 3 | 2.3320 | $1.1700 \times 10^{-1}$ |
| 35 | 11 | 16 | 5 | 1.1074 | $2.1437 \times 10^{-2}$ |
| 66 | 10 | 24 | 8 | $3.9332 \times 10^{-1}$ | $2.8304 \times 10^{-3}$ |
| 86 | 30 | 40 | 10 | $2.6836 \times 10^{-1}$ | $1.3069 \times 10^{-3}$ |
| 135 | 29 | 42 | 13 | $1.4577 \times 10^{-1}$ | $3.4114 \times 10^{-4}$ |
| 185 | 26 | 64 | 20 | $8.5667 \times 10^{-2}$ | $1.0860 \times 10^{-4}$ |
| 266 | 27 | 69 | 27 | $5.0586 \times 10^{-2}$ | $3.5702 \times 10^{-5}$ |
| 418 | 90 | 130 | 40 | $2.1688 \times 10^{-2}$ | $6.3870 \times 10^{-6}$ |
| 597 | 63 | 169 | 55 | $1.3007 \times 10^{-2}$ | $2.0000 \times 10^{-6}$ |
| 828 | 285 | 358 | 72 | $7.7157 \times 10^{-3}$ | $7.1265 \times 10^{-7}$ |
| 1, 010 | 140 | 237 | 86 | $5.2751 \times 10^{-3}$ | $2.9203 \times 10^{-7}$ |

**Continued**

| $n$ | $h_2$ | $h_3$ | $\rho(n,\mathbf{h})$ | $W_2(n,\mathbf{h})$ | $W_4(n,\mathbf{h})$ |
|---|---|---|---|---|---|
| 1, 220 | 319 | 510 | 108 | $3.6308\times10^{-3}$ | $1.1552\times10^{-7}$ |
| 1, 459 | 256 | 373 | 114 | $2.9263\times10^{-3}$ | $9.7463\times10^{-8}$ |
| 1, 626 | 572 | 712 | 140 | $2.1506\times10^{-3}$ | $4.4293\times10^{-8}$ |
| 1, 958 | 202 | 696 | 162 | $1.5620\times10^{-3}$ | $2.2093\times10^{-8}$ |
| 2, 440 | 638 | 1, 002 | 216 | $1.0313\times10^{-3}$ | $8.5161\times10^{-9}$ |
| 3, 237 | 456 | 1, 107 | 252 | $7.0670\times10^{-4}$ | $4.9006\times10^{-9}$ |
| 4, 044 | 400 | 1, 054 | 308 | $4.5620\times10^{-4}$ | $1.8752\times10^{-9}$ |
| 5, 037 | 580 | 1, 997 | 390 | $3.3527\times10^{-4}$ | $9.8872\times10^{-10}$ |
| 6, 066 | 600 | 1, 581 | 460 | $2.3416\times10^{-4}$ | $4.6664\times10^{-10}$ |
| 8, 191 | 739 | 5, 515 | 364 | $1.7\times10^{-4}$ | $4.0\times10^{-10}$ |
| 10, 007 | 544 | 5, 733 | 400 | $1.3\times10^{-4}$ | $2.5\times10^{-10}$ |
| 20, 039 | 5, 704 | 12, 319 | 396 | $6.4\times10^{-5}$ | |
| 28, 117 | 19, 449 | 5, 600 | 585 | $3.0\times10^{-5}$ | |
| 39, 029 | 10, 607 | 26, 871 | 570 | $2.1\times10^{-5}$ | |
| 57, 091 | 48, 188 | 21, 101 | 1, 084 | $9.8\times10^{-6}$ | |
| 82, 001 | 21, 252 | 67, 997 | 1.978 | $4.1\times10^{-6}$ | |
| *140, 052 | 34, 590 | 112, 313 | | $3.33\times10^{-6}$ | |
| *314, 694 | 77, 723 | 252, 365 | | $1.23\times10^{-6}$ | |

**3** ($s=4, h_1=1$)

| $n$ | $h_2$ | $h_3$ | $h_4$ | $\rho(n,\mathbf{h})$ | $W_2(n,\mathbf{h})$ | $W_4(n,\mathbf{h})$ |
|---|---|---|---|---|---|---|
| 60 | 8 | 18 | 22 | 4 | 3.3875 | $8.5025\times10^{-2}$ |
| 118 | 18 | 40 | 52 | 6 | 1.4214 | $1.5513\times10^{-2}$ |
| 180 | 8 | 46 | 74 | 8 | $8.1807\times10^{-1}$ | $5.2230\times10^{-3}$ |
| 286 | 16 | 94 | 138 | 12 | $4.4143\times10^{-1}$ | $1.5466\times10^{-3}$ |
| 440 | 21 | 136 | 216 | 15 | $2.5001\times10^{-1}$ | $4.3550\times10^{-4}$ |
| 562 | 53 | 89 | 221 | 20 | $1.8208\times10^{-1}$ | $2.0716\times10^{-4}$ |
| 732 | 248 | 294 | 324 | 24 | $1.1232\times10^{-1}$ | $8.5499\times10^{-5}$ |
| 932 | 116 | 288 | 314 | 26 | $8.0987\times10^{-2}$ | $4.7288\times10^{-5}$ |
| 1, 142 | 150 | 187 | 274 | 32 | $6.7770\times10^{-2}$ | $2.9213\times10^{-5}$ |
| 1, 354 | 492 | 550 | 658 | 40 | $4.5581\times10^{-2}$ | $1.2280\times10^{-5}$ |
| 2, 129 | 766 | 1, 281 | 1, 906 | 32 | $2.7\times10^{-2}$ | |
| 3, 001 | 174 | 266 | 1, 269 | 46 | $1.7\times10^{-2}$ | |
| 4, 001 | 113 | 766 | 2, 537 | 51 | $1.1\times10^{-2}$ | |

**Continued**

| $n$ | $h_2$ | $h_3$ | $h_4$ | $\rho(n, \mathbf{h})$ | $W_2(n, \mathbf{h})$ | $W_4(n, \mathbf{h})$ |
|---|---|---|---|---|---|---|
| 5, 003 | 792 | 1, 889 | 191 | 32 | $9.2 \times 10^{-3}$ | |
| 6, 007 | 1.351 | 5, 080 | 3, 086 | 80 | $5.9 \times 10^{-3}$ | |
| 8, 191 | 2.488 | 5, 939 | 7, 859 | 72 | $3.8 \times 10^{-3}$ | |
| 10, 007 | 1.206 | 3, 421 | 2, 842 | 84 | $3.0 \times 10^{-3}$ | |
| 20, 039 | 19, 668 | 17, 407 | 14, 600 | 60 | $1.6 \times 10^{-3}$ | |
| 28, 117 | 17, 549 | 1, 900 | 24, 455 | 144 | $6.5 \times 10^{-4}$ | |
| 39, 029 | 30, 699 | 34, 367 | 605 | 135 | $4.9 \times 10^{-4}$ | |
| 57, 091 | 52, 590 | 48, 787 | 38, 790 | 268 | $2.8 \times 10^{-4}$ | |
| 82, 001 | 57, 270 | 58, 903 | 17, 672 | 260 | $1.7 \times 10^{-4}$ | |
| 100, 063 | 92, 313 | 24, 700 | 95, 582 | 352 | $1.1 \times 10^{-4}$ | |
| *147, 312 | 136, 641 | 116, 072 | 76, 424 | | $8.5376 \times 10^{-5}$ | |

**4** $(s = 5, h_1 = 1)$

| $n$ | $h_2$ | $h_3$ | $h_4$ | $h_5$ | $\rho(n, \mathbf{h})$ | $W_2(n, \mathbf{h})$ | $W_4(n, \mathbf{h})$ |
|---|---|---|---|---|---|---|---|
| 1, 069 | 63 | 762 | 970 | 177 | 6 | $7.4 \times 10^{-1}$ | $4.7 \times 10^{-3}$ |
| 1, 543 | 58 | 278 | 694 | 134 | 8 | $4.2 \times 10^{-1}$ | $1.5 \times 10^{-3}$ |
| 2, 129 | 618 | 833 | 1, 705 | 1, 964 | 9 | $3.1 \times 10^{-1}$ | $1.1 \times 10^{-3}$ |
| 3, 001 | 408 | 1, 409 | 1, 681 | 1, 620 | 18 | $1.7 \times 10^{-1}$ | $1.3 \times 10^{-4}$ |
| 4, 001 | 1, 534 | 568 | 3, 095 | 2, 544 | 17 | $1.2 \times 10^{-1}$ | $1.1 \times 10^{-4}$ |
| 5, 003 | 840 | 177 | 3, 593 | 1, 311 | 16 | $9.2 \times 10^{-2}$ | $7.6 \times 10^{-5}$ |
| 6, 007 | 509 | 780 | 558 | 1, 693 | 22 | $7.0 \times 10^{-2}$ | $3.5 \times 10^{-5}$ |
| 8, 191 | 1, 386 | 4, 302 | 7, 715 | 3, 735 | 30 | $4.3 \times 10^{-2}$ | $1.0 \times 10^{-5}$ |
| 10, 007 | 198 | 9, 183 | 6, 967 | 8, 507 | 36 | $3.4 \times 10^{-2}$ | $7.2 \times 10^{-6}$ |
| 15, 019 | 10, 641 | 2, 640 | 6, 710 | 784 | 18 | $2.9 \times 10^{-2}$ | $2.5 \times 10^{-5}$ |
| 20, 039 | 11, 327 | 11, 251 | 12, 076 | 18, 677 | 21 | $1.8 \times 10^{-2}$ | $1.2 \times 10^{-5}$ |
| 33, 139 | 32, 133 | 17, 866 | 21, 281 | 32, 247 | 60 | $8.5 \times 10^{-3}$ | $8.9 \times 10^{-7}$ |
| 51, 097 | 44, 672 | 45, 346 | 7, 044 | 14, 242 | 35 | $5.4 \times 10^{-3}$ | $1.5 \times 10^{-6}$ |
| 71, 053 | 33, 755 | 65, 170 | 12, 740 | 6, 878 | 80 | $2.8 \times 10^{-3}$ | $1.1 \times 10^{-5}$ |
| 100, 063 | 90, 036 | 77, 477 | 27, 253 | 6, 222 | 96 | $1.7 \times 10^{-3}$ | $5.8 \times 10^{-8}$ |
| *374, 181 | 343, 867 | 255, 381 | 310, 881 | 115, 892 | | $1.01 \times 10^{-3}$ | |

**5** $(s = 6, h_1 = 1)$

| $n$ | $h_2$ | $h_3$ | $h_4$ | $h_5$ | $h_6$ | $\rho(n, \mathbf{h})$ | $W_2(n, \mathbf{h})$ | $W_4(n, \mathbf{h})$ |
|---|---|---|---|---|---|---|---|---|
| 2, 129 | 41 | 1, 681 | 793 | 578 | 279 | 4 | 2.0 | $1.9 \times 10^{-2}$ |
| 3, 001 | 233 | 271 | 122 | 1, 417 | 51 | 8 | 1.3 | $5.9 \times 10^{-3}$ |

**Continued**

| $n$ | $h_2$ | $h_3$ | $h_4$ | $h_5$ | $h_6$ | $\rho(n, \mathbf{h})$ | $W_2(n, \mathbf{h})$ | $W_4(n, \mathbf{h})$ |
|---|---|---|---|---|---|---|---|---|
| 4, 001 | 1, 751 | 1, 235 | 1, 945 | 844 | 1, 475 | 6 | $9.5 \times 10^{-1}$ | $4.4 \times 10^{-3}$ |
| 5, 003 | 2, 037 | 1, 882 | 1, 336 | 4, 803 | 2, 846 | 8 | $6.8 \times 10^{-1}$ | $1.6 \times 10^{-3}$ |
| 6, 007 | 312 | 1, 232 | 5, 943 | 4, 060 | 5, 250 | 9 | $5.6 \times 10^{-1}$ | $1.1 \times 10^{-3}$ |
| 8, 191 | 1, 632 | 1, 349 | 6, 380 | 1, 399 | 6, 070 | 12 | $3.7 \times 10^{-1}$ | $5.0 \times 10^{-4}$ |
| 10, 007 | 2, 240 | 4, 093 | 1, 908 | 931 | 3, 984 | 12 | $2.9 \times 10^{-1}$ | $3.8 \times 10^{-4}$ |
| 15, 019 | 8, 743 | 8, 358 | 6, 559 | 2, 795 | 772 | 8 | $2.0 \times 10^{-1}$ | $6.9 \times 10^{-4}$ |
| 20, 039 | 5, 557 | 150 | 11, 951 | 2, 461 | 9, 179 | 12 | $1.3 \times 10^{-1}$ | $1.7 \times 10^{-4}$ |
| 33, 139 | 18, 236 | 1, 831 | 19, 143 | 5, 522 | 22, 910 | 18 | $6.8 \times 10^{-2}$ | $3.5 \times 10^{-5}$ |
| 51, 097 | 9, 931 | 7, 551 | 29, 682 | 44, 446 | 17, 340 | 24 | $4.2 \times 10^{-2}$ | $1.8 \times 10^{-5}$ |
| 71, 053 | 18, 010 | 3, 155 | 50, 203 | 6, 605 | 13, 328 | 18 | $3.3 \times 10^{-2}$ | $2.5 \times 10^{-5}$ |
| 100, 063 | 43, 307 | 15, 440 | 39, 114 | 43, 534 | 39, 955 | 30 | $1.8 \times 10^{-2}$ | $4.5 \times 10^{-6}$ |
| *114, 174 | 107, 538 | 88, 018 | 15, 543 | 80, 974 | 56, 747 | | $1.47 \times 10^{-2}$ | |
| *302, 686 | 285, 095 | 233, 344 | 41, 204 | 214, 668 | 150, 441 | | $4.06 \times 10^{-3}$ | |

**6** $(s = 7, h_1 = 1)$

| $n$ | $h_2$ | $h_3$ | $h_4$ | $h_5$ | $h_6$ | $h_7$ | $\rho(n, \mathbf{h})$ | $W_2(n, \mathbf{h})$ | $W_4(n, \mathbf{h})$ |
|---|---|---|---|---|---|---|---|---|---|
| *3, 997 | 3, 888 | 3, 564 | 3, 034 | 2, 311 | 1, 417 | 375 | | 5.8 | $1.1 \times 10^{-1}$ |
| *11, 215 | 10, 909 | 10, 000 | 8, 512 | 6, 485 | 3, 976 | 1, 053 | | 1.9 | $2.2 \times 10^{-2}$ |
| 15, 019 | 12, 439 | 2, 983 | 8, 607 | 7, 041 | 7, 210 | 6.741 | 6 | 1.2 | |
| 24, 041 | 1, 833 | 18, 190 | 21, 444 | 23, 858 | 1, 135 | 12, 929 | 6 | $6.9 \times 10^{-1}$ | $2.5 \times 10^{-3}$ |
| 33, 139 | 7, 642 | 9, 246 | 5, 584 | 23, 035 | 32, 214 | 30, 396 | 6 | $5.0 \times 10^{-1}$ | $1.9 \times 10^{-3}$ |
| 46, 213 | 37, 900 | 17, 534 | 41, 873 | 32, 280 | 15, 251 | 26, 909 | 12 | $3.3 \times 10^{-1}$ | $3.5 \times 10^{-4}$ |
| 57, 091 | 35, 571 | 45, 299 | 51, 436 | 34, 679 | 1, 472 | 8, 065 | 12 | $2.5 \times 10^{-1}$ | $3.0 \times 10^{-4}$ |
| 71, 053 | 31, 874 | 36, 082 | 13, 810 | 6, 605 | 68, 784 | 9, 848 | 10 | $2.1 \times 10^{-1}$ | $4.5 \times 10^{-4}$ |
| *84, 523 | 82, 217 | 75, 364 | 64, 149 | 48, 878 | 29, 969 | 7, 936 | | $2.0 \times 10^{-1}$ | $6.2 \times 10^{-4}$ |
| 100, 063 | 39, 040 | 62, 047 | 89, 839 | 6, 347 | 30, 892 | 64, 404 | 16 | $1.4 \times 10^{-1}$ | $1.4 \times 10^{-4}$ |
| *172, 155 | 167, 459 | 153, 499 | 130, 657 | 99, 554 | 61, 040 | 18, 165 | | $7.3 \times 10^{-2}$ | $5.4 \times 10^{-5}$ |
| *234, 646 | 228, 245 | 209, 218 | 178, 084 | 135, 691 | 83, 197 | 22, 032 | | $8.0 \times 10^{-2}$ | $4.6 \times 10^{-5}$ |
| *462, 891 | 450, 265 | 412, 730 | 351, 310 | 267, 681 | 164, 124 | 43, 464 | | $1.9 \times 10^{-2}$ | $3.4 \times 10^{-6}$ |
| *769, 518 | 748, 528 | 686, 129 | 584, 024 | 444, 998 | 272, 843 | 72, 255 | | $1.2 \times 10^{-2}$ | |
| *957, 838 | 931, 711 | 854, 041 | 726, 949 | 553, 900 | 339, 614 | 89, 937 | | $8.0 \times 10^{-3}$ | |

7 ($s = 8, h_1 = 1$)

| $n$ | $h_2$ | $h_3$ | $h_4$ | $h_5$ | $h_6$ | $h_7$ | $h_8$ | $\rho(n, \mathbf{h})$ | $W_2(n, \mathbf{h})$ | $W_4(n, \mathbf{h})$ |
|---|---|---|---|---|---|---|---|---|---|---|
| *3, 997 | 3, 888 | 3, 564 | 3, 034 | 2, 311 | 1, 417 | 375 | 3, 211 | | $2.8 \times 10$ | 2.7 |
| *11, 215 | 10, 909 | 10, 000 | 8, 512 | 6, 485 | 3, 976 | 1, 053 | 9, 010 | | 9.6 | $3.5 \times 10^{-1}$ |
| 24, 041 | 17, 441 | 21, 749 | 5, 411 | 12, 326 | 3, 144 | 21, 024 | 6, 252 | 3 | 3.9 | $4.4 \times 10^{-2}$ |
| *28, 832 | 27, 850 | 24, 938 | 20, 195 | 13, 782 | 5, 918 | 25, 703 | 15, 781 | | 3.5 | $4.2 \times 10^{-2}$ |
| 33, 139 | 3, 520 | 29, 553 | 3, 239 | 1, 464 | 16, 735 | 19, 197 | 3, 019 | 6 | 2.7 | $1.2 \times 10^{-2}$ |
| 46, 213 | 5, 347 | 30, 775 | 35, 645 | 11, 403 | 16, 894 | 32, 016 | 16, 600 | 4 | 1.9 | $1.4 \times 10^{-2}$ |
| 57, 091 | 17, 411 | 46, 802 | 9, 779 | 16, 807 | 35, 302 | 1, 416 | 47, 755 | 6 | 1.5 | $4.6 \times 10^{-3}$ |
| 71, 053 | 60, 759 | 26, 413 | 24, 409 | 48, 215 | 51, 048 | 19, 876 | 29, 096 | 6 | 1.2 | $4.2 \times 10^{-3}$ |
| *84, 523 | 82, 217 | 75, 364 | 64, 149 | 48, 878 | 29, 969 | 7, 936 | 67, 905 | | $9.9 \times 10^{-1}$ | $2.2 \times 10^{-3}$ |
| 100, 063 | 4, 344 | 58, 492 | 29, 291 | 60, 031 | 10, 486 | 22, 519 | 60, 985 | 9 | $7.6 \times 10^{-1}$ | $1.0 \times 10^{-3}$ |
| *172, 155 | 167, 459 | 153, 499 | 130, 657 | 99, 554 | 61, 040 | 18, 165 | 138, 308 | | $4.6 \times 10^{-1}$ | $1.2 \times 10^{-3}$ |
| *234, 646 | 228, 245 | 209, 218 | 178, 084 | 135, 691 | 83, 197 | 22, 032 | 188, 512 | | $4.1 \times 10^{-1}$ | $1.3 \times 10^{-3}$ |
| *462, 891 | 450, 265 | 412, 730 | 351, 310 | 267, 681 | 164, 124 | 43, 464 | 371, 882 | | $1.6 \times 10^{-1}$ | $2.2 \times 10^{-4}$ |
| *769, 518 | 748, 528 | 686, 129 | 584, 024 | 444, 998 | 272, 843 | 72, 255 | 618, 224 | | $1.2 \times 10^{-1}$ | $5.3 \times 10^{-4}$ |
| *957, 838 | 931, 711 | 854, 041 | 726, 949 | 553, 900 | 339, 614 | 89, 937 | 769, 518 | | $8.0 \times 10^{-2}$ | $1.4 \times 10^{-4}$ |

8 ($s = 9, h_1 = 1$)

| $n$ | $h_2$ | $h_3$ | $h_4$ | $h_5$ | $h_6$ | $h_7$ | $h_8$ | $h_9$ | $\rho(n, \mathbf{h})$ | $W_2(n, \mathbf{h})$ | $W_4(n, \mathbf{h})$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *3, 997 | 3, 888 | 3, 564 | 3, 034 | 2, 311 | 1, 417 | 375 | 3, 211 | 1, 962 | | $1.2 \times 10^2$ | 6.7 |
| *11, 215 | 10, 909 | 10, 000 | 8, 512 | 6, 485 | 3, 976 | 1, 053 | 9, 010 | 5, 506 | | $4.2 \times 10$ | $9.2 \times 10^{-1}$ |
| 33, 139 | 68 | 4, 624 | 16, 181 | 6, 721 | 26, 221 | 26, 661 | 23, 442 | 3, 384 | 3 | $1.4 \times 10$ | $1.7 \times 10^{-1}$ |
| *42, 570 | 41, 409 | 37, 957 | 32, 308 | 24, 617 | 15, 094 | 3, 997 | 34, 200 | 20.901 | | $1.0 \times 10$ | $1.2 \times 10^{-1}$ |
| 46, 213 | 8, 871 | 40, 115 | 20, 065 | 30, 352 | 15, 654 | 42, 782 | 17, 966 | 33, 962 | 3 | 9.5 | $2.1 \times 10^{-1}$ |
| 57, 091 | 20, 176 | 12, 146 | 23, 124 | 2, 172 | 33, 475 | 5, 070 | 42, 339 | 36, 122 | 4 | 7.5 | $4.6 \times 10^{-2}$ |
| 71, 053 | 26, 454 | 13, 119 | 27, 174 | 17, 795 | 22, 805 | 43, 500 | 45, 665 | 49, 857 | 4 | 6.0 | $4.3 \times 10^{-2}$ |
| 100, 063 | 70, 893 | 53, 211 | 12, 386 | 27, 873 | 56, 528 | 16, 417 | 17, 628 | 14, 997 | 6 | 4.1 | $1.3 \times 10^{-2}$ |
| 159, 053 | 60, 128 | 101, 694 | 23, 300 | 43, 576 | 57, 659 | 42, 111 | 85, 501 | 93, 062 | 8 | 2.5 | $6.6 \times 10^{-3}$ |
| *172, 155 | 167, 459 | 153, 499 | 130, 657 | 99, 554 | 61, 040 | 18, 165 | 138, 308 | 84, 523 | | 2.4 | $7.6 \times 10^{-3}$ |
| *234, 646 | 228, 245 | 209, 218 | 178, 084 | 135, 691 | 83, 197 | 22, 032 | 188, 512 | 115, 204 | | 1.9 | $2.0 \times 10^{-2}$ |
| *462, 891 | 450, 265 | 412, 730 | 351, 310 | 267, 681 | 164, 124 | 43, 464 | 371, 882 | 227, 266 | | $9.8 \times 10^{-1}$ | $1.1 \times 10^{-2}$ |
| *769, 518 | 748, 528 | 686, 129 | 584, 024 | 444, 998 | 272, 843 | 72, 255 | 618, 224 | 377, 811 | | $5.3 \times 10^{-1}$ | $1.4 \times 10^{-3}$ |
| *957, 838 | 931, 711 | 854, 041 | 726, 949 | 553, 900 | 339, 614 | 89, 937 | 769, 518 | 470, 271 | | $4.1 \times 10^{-1}$ | $1.9 \times 10^{-3}$ |

9 $(s = 10, h_1 = 1)$

| $n$ | $h_2$ | $h_3$ | $h_4$ | $h_5$ | $h_6$ | $h_7$ | $h_8$ | $h_9$ | $h_{10}$ | $\rho(n, \mathbf{h})$ | $W_2(n, \mathbf{h})$ | $W_4(n, \mathbf{h})$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *4, 661 | 4, 574 | 4, 315 | 3, 889 | 3, 304 | 2, 570 | 1, 702 | 715 | 4, 289 | 3, 122 | | $4.5 \times 10^2$ | $2.2 \times 10$ |
| *13, 587 | 13, 334 | 12, 579 | 11, 337 | 9, 631 | 7, 492 | 4, 961 | 2, 084 | 12, 502 | 9, 100 | | $1.6 \times 10^2$ | 8.7 |
| *24, 076 | 23, 628 | 22, 290 | 20, 090 | 17, 066 | 13, 276 | 8, 790 | 3, 692 | 22, 153 | 16, 125 | | $8.8 \times 10$ | 5.1 |
| *58, 358 | 57, 271 | 54, 030 | 48, 695 | 41, 366 | 32, 180 | 21, 307 | 8, 950 | 53, 697 | 39, 086 | | $3.6 \times 10$ | 2.4 |
| 85, 633 | 37, 677 | 35, 345 | 3, 864 | 54, 821 | 74, 078 | 30, 354 | 57, 935 | 51, 906 | 56, 279 | 2 | $2.4 \times 10$ | 2.3 |
| 103, 661 | 45, 681 | 57, 831 | 80, 987 | 9, 718 | 51, 556 | 55, 377 | 37, 354 | 4, 353 | 27, 595 | 2 | $2.1 \times 10$ | $4.2 \times 10^{-1}$ |
| 115, 069 | 65, 470 | 650 | 95, 039 | 77, 293 | 98, 366 | 70, 366 | 74, 605 | 55, 507 | 49, 201 | 2 | $1.7 \times 10$ | $3.0 \times 10^{-1}$ |
| 130, 703 | 64, 709 | 53, 373 | 17, 385 | 5, 244 | 29, 008 | 52, 889 | 66, 949 | 51, 906 | 110, 363 | 4 | $1.4 \times 10$ | $8.9 \times 10^{-2}$ |
| 155, 093 | 90, 485 | 20, 662 | 110, 048 | 102, 308 | 148, 396 | 125, 399 | 124, 635 | 10, 480 | 44, 198 | 4 | $1.2 \times 10$ | $6.9 \times 10^{-2}$ |
| *805, 098 | 790, 101 | 745, 388 | 671, 792 | 570, 685 | 443, 949 | 293, 946 | 123, 470 | 740, 795 | 539, 222 | | 2.3 | $2.8 \times 10^{-2}$ |

10 $(s = 11, h_1 = 1)^*$

| $n$ | 4, 661 | 13, 587 | 24, 076 | 58, 358 | 297, 974 | 698, 047 | 1, 243, 423 | 2, 226, 963 | 7, 494, 007 |
|---|---|---|---|---|---|---|---|---|---|
| $h_2$ | 4, 574 | 13, 334 | 23, 628 | 57, 271 | 294, 481 | 685, 041 | 1, 228, 845 | 2, 200, 854 | 7, 354, 408 |
| $h_3$ | 4, 315 | 12, 579 | 22, 290 | 54, 030 | 284, 041 | 646, 274 | 1, 185, 282 | 2, 122, 833 | 6, 938, 211 |
| $h_4$ | 3, 889 | 11, 337 | 20, 090 | 48, 695 | 266, 778 | 582, 461 | 1, 113, 244 | 1, 993, 814 | 6, 253, 169 |
| $h_5$ | 3, 304 | 9, 631 | 17, 066 | 41, 366 | 242, 894 | 494, 796 | 1, 013, 577 | 1, 815, 311 | 5, 312, 043 |
| $h_6$ | 2, 570 | 7, 492 | 13, 276 | 32, 180 | 212, 668 | 384, 914 | 887, 449 | 1, 589, 415 | 4, 132, 365 |
| $h_7$ | 1, 702 | 4, 961 | 8, 790 | 21, 307 | 176, 456 | 254, 860 | 736, 338 | 1, 318, 777 | 2, 736, 109 |
| $h_8$ | 715 | 2, 084 | 3, 692 | 8, 950 | 134, 682 | 107, 051 | 562, 016 | 1, 006, 567 | 1, 149, 286 |
| $h_9$ | 4, 289 | 12, 502 | 22, 153 | 53, 697 | 87, 835 | 642, 292 | 366, 527 | 656, 448 | 6, 895, 461 |
| $h_{10}$ | 3, 122 | 9, 100 | 16, 125 | 39, 086 | 36, 464 | 467, 527 | 152, 163 | 272, 523 | 5, 019, 180 |
| $h_{11}$ | 1, 897 | 5, 529 | 9, 797 | 23, 747 | 297, 147 | 284, 044 | 1, 164, 860 | 2, 086, 257 | 3, 049, 402 |
| $W_2(n, \mathbf{h})$ | $1.9 \times 10^3$ | $6.6 \times 10^2$ | $3.7 \times 10^2$ | $1.5 \times 10^2$ | $3.1 \times 10$ | $1.2 \times 10$ | 8.1 | 3.6 | $6.4 \times 10^{-1}$ |
| $W_4(n, \mathbf{h})$ | $7.7 \times 10$ | $2.0 \times 10$ | $1.1 \times 10$ | 4.2 | $7.0 \times 10^{-1}$ | $7.6 \times 10^{-2}$ | | | |

11 $(s = 12, 13, 14, h_1 = 1)^*$

| $n$ | 18, 984 | 53, 328 | 77, 431 | 297, 974 | 1, 243, 423 | 2, 428, 705 | 14, 753, 436 | 19, 984, 698 | 34, 248, 063 |
|---|---|---|---|---|---|---|---|---|---|
| $h_2$ | 18, 761 | 52, 703 | 76, 523 | 294, 481 | 1, 228, 845 | 2, 400, 231 | 14, 580, 465 | 19, 750, 396 | 33, 846, 536 |
| $h_3$ | 18, 096 | 50, 834 | 73, 810 | 284, 041 | 1, 185, 282 | 2, 315, 141 | 14, 063, 582 | 19, 050, 236 | 32, 646, 662 |
| $h_4$ | 16, 996 | 47, 745 | 69, 324 | 266, 778 | 1, 113, 244 | 2, 174, 435 | 13, 208, 845 | 17, 892, 427 | 30, 662, 508 |
| $h_5$ | 15, 475 | 43, 470 | 63, 118 | 242, 894 | 1, 013, 577 | 1, 979, 761 | 12, 026, 276 | 16, 290, 543 | 27, 917, 337 |
| $h_6$ | 13, 549 | 38, 061 | 55, 264 | 212, 668 | 887, 449 | 1, 733, 402 | 10, 529, 739 | 14, 263, 366 | 24, 443, 334 |
| $h_7$ | 11, 242 | 31, 580 | 45, 854 | 176, 456 | 736, 338 | 1, 438, 245 | 8, 736, 780 | 11, 834, 661 | 20, 281, 228 |
| $h_8$ | 8, 581 | 24, 104 | 34, 998 | 134, 682 | 562, 016 | 1, 097, 753 | 6, 668, 420 | 9, 032, 903 | 15, 479, 816 |
| $h_9$ | 5, 596 | 15, 720 | 22, 825 | 87, 835 | 366, 527 | 715, 916 | 4, 348, 908 | 5, 890, 941 | 10, 095, 390 |
| $h_{10}$ | 2, 323 | 6, 526 | 9, 476 | 36, 464 | 152, 163 | 297, 211 | 1, 805, 439 | 2, 445, 610 | 4, 191, 077 |
| $h_{11}$ | 17, 785 | 49, 959 | 72, 539 | 279, 147 | 1, 164, 860 | 2, 275, 252 | 13, 821, 268 | 18, 722, 002 | 32, 084, 164 |
| $h_{12}$ | 14, 053 | 39, 477 | 57, 320 | 220, 583 | 920, 477 | 1, 797, 913 | 10, 921, 619 | 14, 794, 199 | 25, 353, 030 |
| $h_{13}$ | 10, 158 | 28, 534 | 41, 430 | 159, 433 | 665, 302 | 1, 299, 495 | 7, 893, 924 | 10, 692, 946 | 18, 324, 655 |
| $h_{14}$ | 6, 143 | 17, 255 | 25, 054 | 96, 414 | 402, 327 | 785, 841 | 4, 773, 681 | 6, 466, 329 | 11, 081, 440 |
| $W_2(12, n, \mathbf{h})$ | $1.3 \times 10^3$ | $7.2 \times 10^2$ | $5.1 \times 10^2$ | $1.3 \times 10^2$ | $3.1 \times 10$ | $1.6 \times 10$ | | | |
| $W_4(12, n, \mathbf{h})$ | $6.9 \times 10$ | $1.5 \times 10$ | $1.3 \times 10$ | 1.9 | $4.2 \times 10^{-1}$ | | | | |
| $W_2(13, n, \mathbf{h})$ | $5.7 \times 10^3$ | $3.1 \times 10^3$ | $2.1 \times 10^3$ | $5.6 \times 10^2$ | $1.3 \times 10^2$ | $6.9 \times 10$ | $1.0 \times 10$ | 8.8 | 4.0 |
| $W_4(13, n, \mathbf{h})$ | $1.5 \times 10^2$ | $3.7 \times 10$ | $2.6 \times 10$ | 5.9 | 1.8 | $3.6 \times 10^{-1}$ | | | |
| $W_2(14, n, \mathbf{h})$ | $3.8 \times 10^4$ | $1.3 \times 10^4$ | $9.2 \times 10^3$ | $2.4 \times 10^3$ | $5.8 \times 10^2$ | $3.0 \times 10^2$ | $4.7 \times 10$ | $3.5 \times 10$ | $2.0 \times 10$ |
| $W_4(14, n, \mathbf{h})$ | $5.1 \times 10^2$ | $1.8 \times 10^2$ | $1.4 \times 10^2$ | $3.6 \times 10$ | $1.1 \times 10$ | 5.4 | | | |

**12** $(s = 15, 16, 17, 18, h_1 = 1)^*$

| $n$ | 70, 864 | 139, 489 | 1, 139, 691 | 2, 422, 957 | 4, 395, 774 | 14, 271, 038 | 55, 879, 244 |
|---|---|---|---|---|---|---|---|
| $h_2$ | 70, 353 | 138, 484 | 1, 131, 480 | 2, 398, 094 | 4, 364, 102 | 14, 168, 215 | 55, 476, 633 |
| $h_3$ | 68, 825 | 135, 476 | 1, 106, 904 | 2, 323, 761 | 4, 269, 316 | 13, 860, 486 | 54, 271, 700 |
| $h_4$ | 66, 291 | 130, 487 | 1, 066, 142 | 2, 200, 720 | 4, 112, 097 | 13, 350, 069 | 52, 273, 127 |
| $h_5$ | 62, 768 | 123, 553 | 1, 009, 487 | 2, 030, 234 | 3, 893, 578 | 12, 640, 642 | 49, 495, 314 |
| $h_6$ | 58, 283 | 114, 724 | 937, 347 | 1, 814, 052 | 3, 615, 335 | 11, 737, 315 | 45, 958, 274 |
| $h_7$ | 52, 867 | 104, 063 | 850, 242 | 1, 554, 392 | 3, 279, 371 | 10, 646, 597 | 41, 687, 793 |
| $h_8$ | 46, 559 | 91, 647 | 748, 799 | 1, 253, 920 | 2, 888, 108 | 9, 376, 347 | 36, 713, 742 |
| $h_9$ | 39, 405 | 77, 566 | 633, 750 | 915, 717 | 2, 444, 365 | 7, 935, 718 | 31, 072, 856 |
| $h_{10}$ | 31, 457 | 61, 921 | 505, 923 | 543, 256 | 1, 951, 338 | 6, 335, 088 | 24, 805, 477 |
| $h_{11}$ | 22, 772 | 44, 825 | 366, 239 | 140, 357 | 1, 412, 580 | 4, 585, 990 | 17, 956, 764 |
| $h_{12}$ | 13, 412 | 26, 401 | 215, 705 | 2, 134, 112 | 831, 972 | 2, 701, 027 | 10, 576, 061 |
| $h_{13}$ | 3, 445 | 6, 781 | 55, 406 | 1, 683, 011 | 213, 699 | 693, 780 | 50, 314, 090 |
| $h_{14}$ | 63, 806 | 125, 597 | 1, 026, 186 | 1, 214, 641 | 3, 957, 988 | 12, 849, 750 | 41, 669, 876 |
| $h_{15}$ | 52, 844 | 104, 019 | 849, 882 | 733, 806 | 3, 277, 986 | 10, 642, 098 | 32, 725, 430 |
| $h_{16}$ | 41, 501 | 81, 691 | 667, 455 | | 2, 574 365 | 8, 357, 770 | 23, 545, 197 |
| $h_{17}$ | 29, 859 | 58, 775 | 480, 219 | | 1, 852, 197 | 6, 013, 224 | 14, 195, 319 |
| $h_{18}$ | 18, 002 | 35, 435 | 289, 522 | | 1, 116, 683 | 3, 625, 352 | 2, 716, 545 |
| $W_2(15, n, \mathbf{h})$ | $4.3 \times 10^4$ | $2.2 \times 10^4$ | $2.7 \times 10^3$ | $1.3 \times 10^3$ | $7.0 \times 10^2$ | $2.2 \times 10^2$ | $5.5 \times 10$ |
| $W_4(15, n, \mathbf{h})$ | $4.6 \times 10^2$ | $2.4 \times 10^2$ | $2.4 \times 10$ | $1.2 \times 10$ | 6.8 | 2.8 | |
| $W_2(16, n, \mathbf{h})$ | $1.9 \times 10^5$ | $9.4 \times 10^4$ | $1.2 \times 10^4$ | | $3.0 \times 10^3$ | $9.2 \times 10^2$ | $2.4 \times 10^2$ |
| $W_4(16, n, \mathbf{h})$ | $1.5 \times 10^3$ | $7.5 \times 10^2$ | $8.1 \times 10$ | | $2.3 \times 10$ | 7.2 | $7.2 \times 10^{-1}$ |
| $W_2(17, n, \mathbf{h})$ | $8.0 \times 10^5$ | $4.0 \times 10^5$ | $5.0 \times 10^4$ | | $1.3 \times 10^4$ | $4.0 \times 10^3$ | $1.0 \times 10^3$ |
| $W_4(17, n, \mathbf{h})$ | $4.5 \times 10^3$ | $2.3 \times 10^3$ | $2.7 \times 10^2$ | | $7.0 \times 10$ | $2.3 \times 10$ | 4.1 |
| $W_2(18, n, \mathbf{h})$ | $3.4 \times 10^6$ | $1.7 \times 10^6$ | $2.1 \times 10^5$ | | $5.5 \times 10^4$ | $1.7 \times 10^4$ | $4.3 \times 10^3$ |
| $W_4(18, n, \mathbf{h})$ | $1.4 \times 10^4$ | $7.3 \times 10^3$ | $8.7 \times 10^2$ | | $2.3 \times 10^2$ | $7.1 \times 10$ | |

# Bibliography

C. R. Adams, and J. A. Clarkson,

[1] On definitions of bounded variation for function of two variables, *Trans. Amer. Math. Soc.*, **35**, 1933, 824–854.

[2] Properties of functions $f(x,y)$ of bounded variation, *Trans. Amer. Math. Soc.*, **36**, 1934, 711–730.

N. S. Bahvalov,

[1] Approximate computation of multiple integrals, *Vestnik Moskow Univ., Ser. Mat. Meh. Astr. Fiz. Him.*, **4**, 1959, 3–18.

[2] An estimate of the main remainder term in quadrature formula, *Z. Vycisl. Mat. i Mat. Fiz.*, **1**, 1961, 64–77.

[3] On embedding theorems for class of functions with bounded derivatives, *Vestnik Moskow Univ., Ser Mat. Meh. Astr. Fiz. Him.*, **3**, 1963, 7–16.

[4] Optimal convergence bounds for quadrature processes and integration methods of Monte Carlo type for classes of functions, *Z. Vycisl. Mat. i Mat. Fiz.*, **4**, suppl., 1964, 5–63.

A. Baker,

[1] On some Diophantine inequalities involving the exponential function, *Canad. J. Math.*, **17**, 1965, 616–626.

L. Bernstein,

[1] The Jacobi-Perron algorism, its theory and applications, Lec. Not. in Math,. Springer Verlag, 207, 1971.

J. W. S. Cassels,

[1] An introduction to Diophantine approximation, Camb. Univ. Press, 1957.

H. Conroy,

[1] Molecular Schrödinger equation, VIII: A new method for the evaluation of multidimensional integrals, *J. Chemical Phys.*, **47**, 1967, 5307–5318.

R. Cranley and T. N. L. Patterson,

[1] Randomization of number theoretic methods for multiple integration, *SIAM J. Numer. Anal.*, **13**, 1976, 904–914.

P. J. Davis and P. Rabinowitz,

[1] Some Monte Carlo experiments in computing multiple integrals, *Math. Tables Aids Comput.*; **10**, 1956, 1–8.

P. Erdös and P. Turán,

[1] On a problem in the theory of uniform distribution I, *Indag. Math.*, **10**, 1948, 370–378.

I. M. Gelfand, A. S. Frolov and N. N. Cencov,

[1] The computation of continuous integrals by the Monte Carlo method, *Izv. Vyss. Ucebn., Zaved. Mat.*, **5**, 1958, 32–45.

S. Haber,

[1] Numerical evaluation of multiple integrals, *SIAM Rev.*; **12**, 1970, 481–526.

[2] Experiments on optimal coefficients, Applications of number theory to numerical analysis (S. K. Zaremba, ed.), Academic Press, New York, 1972, 11–37.

S. Haber and C. F. Osgood,

[1] On the sum $\Sigma\langle na\rangle^{-1}$ and numerical integration, *Pacific J. Math.*, **31**, 1969, 383–394.

J. H. Halton,

[1] On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals, *Numer. Math.*, **2**, 1960, 84–90.

J. M. Hammersley,

[1] Monte Carlo methods for solving multivariable problems, *Ann. New. York Acad. Sci.*, **86**, 1960, 844–874.

G. H. Hardy,

[1] On double Fourier series and especially those which represent the double zeta function with real and incommensurable parameters, *Quart. J. Math.*, Oxford, **37**, 1906, 53–79.

C. B. Haselgrove,

[1] A method for numerical integration, *Math. Comp.*, **15**, 1961, 323–337.

E. Hlawka,

[1] Funktionen von beschränkter Variation in der Theorie Gleichverteilung, *Ann. Mat. pure Appl.*, **54**, 1961, 325–333.

[2] Über die Diskrepanz mehrdimensionaler Folgen mod 1, *Math. Z.*, **77**, 1961, 273–284.

[3] Zur angenäherten Berechnung mehrfacher Integrale, *Monatsh, Math.*, **66**, 1962, 140–151.

[4] Uniform distribution modulo 1 and numerical analysis, *Compositio Math.*, **16**, 1964, 92–105.

[5] Trigonometrische Interpolation bei Funktionen Von mehreren Variablen, *Acta Arith.*, **9**, 1964, 305–320.

Hsu Li Zhi and Zhou Yun Shi,

[1] Numerical evaluation of multiple integrals, Science Press, Beijing, 1980.

Hua Loo Keng,

[1] Corrigendum on a paper of Su Jia Ju concerning the 5-th algebraic equation, *Science*, 2, **15**, 1930, 307.

[2] Introduction to number theory, Science Press, Beijing, 1956.

[3] Starting from "Yang Hui triangle", Qing Nian Press, Beijing, 1956.

[4] Additive prime number theory, Science Press, Beijing, 1957.

[5] The estimation of trigonometrial sum and its application in number theory, Science Press, Beijing, 1963.

Hua Loo Keng and Wang Yuan,

[1] Remarks concerning numerical integration, *Sci. Record*, (N. S.), **4**, 1960, 8–11.

[2] Numerical evaluation of integrals, Science Press, Beijing, 1961.

[3] Numerical integration and its applications, Science Press, Beijing, 1963.

[4] On Diophantine approximations and numerical integrations, (I) *Sci. Sin.*; 6, **13**, 1964, 1007–1008(II) *Sci. Sin.*, 6, **13**, 1964, 1009–1010.

[5] On numerical integration of periodic functions of several variables, *Sci. Sin.*, 7, **14**, 1965, 964–978.

[6] On uniform distribution and numerical analysis (Number theoretic method), (I) *Kexue Tongbao*, **3**, 1973, 112–114, (II) *Kexue Tongbao*, **4**, 1973, 165–166, (III) *Kexue Tongbao*, **12**, 1974, 559–560.

[7] On uniform distribution and numerical analysis (Number theoretic method), (I) *Sci. Sin.*, 4, **16**, 1973, 483–505, (II) *Sci. Sin.*, 3, **17**, 1974, 331–348, (III) *Sci Sin.*, 2, **18**, 1975, 184–198.

[8] A note on simultaneous Diophantine approximations to algebraic integers, *Sci. Sin.*, 5, **20**, 1977, 563–567.

Hua Loo Keng, Wang Yuan and Pei Ding Yi,

[1] On a set of independent units of cyclotomic field, *Ziran Zazhi*, 5, 1978, 6.

P. Keast,

[1] Multi-dimensional quadrature formula, *Tech. Rep.*, **40**, Dept. of Computer Science, Toronto Univ., 1972.

[2] Optimal parameters for multi-dimensional integrals, *SIAM J. on Numer. Anal.*, **10**, 1973, 831–838.

G. Kedem and S. K. Zaremba,

[1] A table of good lattice point in three dimensions, *Numer. Math.*, **23**, 1974, 175–180.

A. Khintchine,

[1] Metrical problems of irrational numbers, *Uspehi Mat. Nauk SSSR*, 1, 1936. 7–37.

I. F. Koksma,

[1] Een algemeene stelling uit de theorie der gelijkmatige Verdeeling modulo 1, *Math.* B (Zutphen), **11**, 1942–1943, 7–11.

[2] Some theorems on Diophantine inequalities, *Math. Cent. Amer.*, Scriptum, **5**, 1950, 1–51.

N. M. Korobov,

[1] Approximate calculation of multiple integrals with the aid of methods in the theory of numbers, *Dokl. Akad. Nauk SSSR*, **115**, 1957, 1062–1065.

[2] The approximate computation of multiple integrals, *Dokl. Akad. Nauk SSSR*, **124**, 1959, 1207–1210.

[3] On the approximate solultion of integral equations, *Dokl. Akad. Nauk SSSR*, **128**, 1959, 233–238.

[4] Computation of multiple integrals by the method of optimal coefficients, *Vestnik Moskow Univ. ser. Mat. Meh. Astr. Fiz. Him.*, **4**, 1959, 19–25.

[5] Properties and calculation of optimal coefficients, *Dokl. Akad. Nauk SSSR*, **132**, 1960, 1009–1012.

[6] Application of number-theoretic nets to integral equations and interpolation formulas, Trudy Mat. Inst. Steklov, **60**, 1961, 195–210.

[7] Number theoretic methods in approximate analysis, Fizmatigiz, Moscow, 1963.

[8] Some problems in the theory of Diophantine approximation, *Uspehi Mat. Nauk SSSR*, 3, **22**, 1967, 73–118.

J. M. Krause,

[1] Fouriersche Reihen mit zwei veranderlichen Grössen, *Ber. Verh. Sächs. Akad. Wiss. Leipzig. Math-naturw. Kl.*; **55**, 1903, 164–197.

L. Kuipers and H. Niederreiter,

[1] Uniform distribution of sequences, Wiley, New York, 1974.

E. Landau,

[1] Vorlesungen über Zahlentheorie, III Chelsea pub. Co., Nuw York, 1947.

J. J. Liang,

[1] On the integral basis of the maximal real subfield of a cyclotomcic field, *J. Reine angew, Math.*, **286/287**, 1976, 223–226.

K. Mahler,

[1] On a paper by A. Baker on the approximation of rational powers of *e*, *Acta Arith.*, **27**, 1975, 61–87.

D. Maisonneuve,

[1] Recherche at utilisation des "bons treillis". Programmation et résultats numériques, Applications of Number Theory to Numerical Analysis (S. K. Zaremba, ed), Academic Press, New York, 1972, 121–201.

J. M. Masley and H. L. Montgomery,

[1] Cyclotomic fields with unique facterization, *J. Reine angew. Math.*, **286/287**, 1976, 248–256,

H. Minkowski,

[1] Über periodische Approximationen algebraischer Zahlen, *Acta Math.*, **26**, 1902, 333-351.

Y. S. Moon,

[1] Some numerical experiments on number-theoretic methods in the approximation of multi-dimensional integrals, *Tech. Rep.*, **72**, Dept. of Computer Science, Toronto Univ., 1974.

H. Niederreiter,

[1] Methods for estimating discrepancy, Applications of Number Theory to Numerical Analysis (S. K. Zaremba, ed.), Academic Press, New York, 1972, 203–236.

[2] Application of Diophantine approximations to numerical integration, Diophantine Approximation and Its Applications (C. F. Osgood, ed.), Academic Press, New York, 1973, 129–199.

[3] Pseudo-random numbers and optimal coefficients, *Advances in Math.*, **26**, 1977, 99–181.

[4] Existence of good lattice points in the sense of Hlawka, *Monatsh. Math.*, **86**, 1978, 203–219.

[5] Quasi-Monte Carlo methods and pseudo-random numbers, *Bull. Amer. Math. Soc.*, 6, **84**, 1978, 957–1041.

L. G. Peck,

[1] On uniform distribution of algebraic numbers, *Proc. Amer. Math. Soc.*, **4**, 1953, 440–443.

O. Perron,

[1] Grundlagen fuer eine Theorie des Jacobische Kettenhruchalgorithmus, *Math. Ann.*, **64**, 1907, 1–76.

C. Picot,

[1] La répartition modulo 1et les nombres algebriques, *Ann. Sc. Norm. Sup. Pisa.*, **2**, 1938, 205–248.

K. Ramachandra,

[1] On the units of cyclotomic fields, *Acta Arith.*, **12**, 1966, 165–173.

G. N. Raney,

[1] Generalization of Fibonacci sequence to $n$ dimensions, *Canad. J. Math.*, **18**, 1966, 332–349.

R. D. Richtmyer,

[1] The evaluation of definite integrals and a quasi-Monte Carlo method based on the properties of algebraic numbers, *Report LA*-1342, *Los Alamos Sci. Lab.*, LosAlamos, N. M. 1951.

V. S. Rjabenkii,

[1] Tables and interpolation of a certain class of functions, *Dokl. Akad. Nauk SSSR*, **131**, 1960, 1025–1027.

[2] A way of obtaining difference schemes and the use of number theoretic nets for the solution of the Caucly problem by the method of finite differences, Trudy Mat. Inst. Steklov, **60**, 1961, 232–237.

K. F. Roth,

[1] On irreqularities distribution, *Mathematika*, **1**, 1954, 73–79.

Yu. N. Sahov,

[1] On the approximate solution of Volterra equation of second type by the method of iteration, *Dokl. Akad. Nauk SSSR*, **128**, 1959, 1136–1139.

[2] On calculating the eigenvalues of a multi-dimensional symmetric kernel using number-theoretic nets, *Z, Vycisl, Mat. i Mat. Fiz.*, **3**, 1963, 988–997.

[3] On the approximate solution of multi-dimensional linear Volterra equation of second type by the method of interation, *Z. Vycisl. Mat. i Mat. Fiz.*, **4**, Suppl., 1964, 75–100.

[4] The calculation of integrals of increasing multiplicity, *Z, Vycisl. Mat. i Mat. Fiz.*, **5**, 1965, 911–916.

A. I. Saltykov,

[1] Tables for computing multiple integrals by the method of optimal coefficients, *Z. Vycisl. Mat. i Mat. Fiz.*, **3**, 1963, 181–186.

I. F. Sarygin,

[1] The use of number-theoretic methods of integration in the case of non-periodic functions, *Dokl. Akad. Nauk. SSSR*, **132**, 1960, 71–74.

[2] A lower estimate for the error of quadrature formulas for certain classes of functions, *Z. Vycisl. Mat. i Mat. Fiz.*, **3**, 1963, 370–376.

W. M. Schmidt,

[1] Metrical theorems on fractional parts of sequences, *Trans. Amer. Math. Soc.*; **110**, 1964, 493–518.

[2] Simultaneous approximation to algebraic numbers by rationals, *Acta Math.*, **125**, 1970, 189–201.

[3] Inregularities of distribution, VII, *Acta Arich.*, **21**, 1972, 45–50.

[4] Diophantine approximation, Lec. Not. in Math. Springer Verlag, 785, 1980.

Shih Shu Chung,

[1] Une generalization de s "bons trellis", *C. R. Acad. Sc. Paris*, 290, 1980, 527–530.

W. Sinnott,

[1] On the stickelberger ideal and the circular units of a cyclotomic field (to appear).

S. A. Smoljak,

[1] Interpolation and quadrature formulas for the classes $W_s^\alpha$ and $E_s^\alpha$, *Dokl. Akad. Nauk SSSR*, **131**, 1960, 1028–1031.

I. M. Sobol,

[1] An exact estimate of the error in multidimensional quadrature formulas for functions of the classes $\tilde{W}_1$ and $\tilde{H}_1$, *Z. Vycisl. Mat. i Mat. Fiz.*, **1**, 1961, 208–216.

V. M. Solodov,

[1] On the calculation of multiple integrals, *Dokl. Akad. Nauk SSSR*, **127**, 1959, 753–756.

[2] Integration over regions different from the unit cube, *Z. Vycisl. Mat. i Mat. Fiz.*, **8**, 1968, 1334–1341.

V. T. Stojancev,

[1] Solution of the Cauchy problem for a parabolic equation by a quasi-Monte-Carlo method, *Z. Vycisl. Mat. i Mat. Fiz.*, **13**, 1973, 1153–1160.

A. H. Stroud,

[1] Approximate calculation of multiple integrals, Prentice-Hall, 1971.

T. Van Aardenne Ehrenfest,

[1] On the imposibility of a just distribution, *Indag. Math.*, **11**, 1949, 264–269.

J. G. Van der Corput,

[1] Verteilungs funktionen, I, *Proc. Akad. Amsterdam*, **38**, 1935, 813–821.

T. Vijayaraghavan,

[1] On the fractional parts of the powers of a number, II, *Proc. Cambridge Phil. Soc.*, **37**, 1941, 349–357.

I. M. Vinogradov,

[1] The method of trigonometrical sums in the theory of numbers, Fizmatgiz, Moscow, 1971.

V. S. Vladimirov,

[1] Equations of mathematical physics, Marcel Dekkel, New York, 1971.

Wang Yuan,

[1] A note on interpolation of a certain class of functions, *Sci, Sin.*, 6, **10**, 1960, 632–636.

[2] On numerical integration and its applications (number-theoretic method), *Shuxue Jinzhan*, 1, **5**, 1962, 1–44.

[3] Remarks on the interpolation of a certain class of functions, *Sci. Sin.*, 4, **14**, 1965, 629–631.

[4] On interpolation of a certain class of functions, *Kexue Tongbao*, 9, 1966, 387–389.

[5] On Diophantine approxination and approximate analysis (to appear)

Wang Yuan and Fang Kai Tai,

[1] A note on uniform distribution and experimental design, *Kexue Tongbao* (to appear).

Wang Yuan, Xu Guang Shan and Zhang Rong Xiao,

[1] On number-theoretic method of numerical integration in multi-dimensional space, I, *Acta Math. Appl. Sin.*, 2, 1978, 106–114, II (to appear).

Wang Yuan, Zhu Yao Cheng and Jian Yun Cui,

[1] Several remarks on number-theoretic method in numerical analysis, *Journal of the Univ. of sci. and tech. of China*, 1965, 213–218.

A. Weil,

[1] On some exponential sums, *Proc. Nat. Acad. Sci.*, USA, **34**, 1948, 204–207.

H. Weyl,

[1] Über die Gleichverteilung der Zahlen mod. Eins, *Math. Ann.*, **77**, 1916, 313–352.

Xie Ting Fan and Pei Ding Yi,

[1] On the irreducibility of polynomials, *Kexue Tongbao*, 9, 1975, 414–415.

Xu Guang Shan,

[1] On the approximate solution of the Cauchy problem of parabolic equation, *Kexue Tongbao*, 8, 1975, 361–364.

S. K. Zaremba,

[1] Good lattice points, discrepancy and numerical integration, *Ann. Mat. Pure Apply.*, **73**, 1996, 293–317.

[2] La méthode des "bons treillis" pour le calcul numérique des intégrales multiples, Applications of number theory to numerical analysis (S. K. Zaremba, ed.), Academic Press, New York, 1972, 39–119.

[3] On Cartesian products of good lattices, *Math. Comp.*, **30**, 1976, 546–552.

Zhang Rong Xiao,

[1] On approximate calculation of multiple integrals (number-theoretic method), *Journal of the Univ. of sci. and tech. of China*, 1964, 76–87.



# （乙）

# 应用统计中的数论方法

# 关于均匀分布与试验设计 (数论方法)*

王元 (中国科学院数学研究所)
方开泰 (中国科学院应用数学研究所)

## §1. 问 题

在一项试验中, 若有 $s$ 个因素, 每个因素各有 $q$ 个水平, 此处 $q>1$, 这种试验常采用正交试验法 [1], 所需试验次数为 $rq^2$, 此处 $r$ 为自然数. 当 $q$ 较大时 (例如 $q\geqslant 9$) 就需要做较多的试验. 因此需要找一种多因素、多水平而试验次数又较少的设计. 为此本文提出均匀设计.

设有 $s$ 个因素, 每个因素各有 $q$ 个水平, 如果所有可能的试验都做, 则共有 $q^s$ 种组合, 正交试验法是从这部分点中选取 $rq^2$ 个点. 均匀设计则是利用数论中一致分布的理论选取 $q$ 个点. 在某些实验应用中表明, 均匀设计比正交设计更为方便.

## §2. 布 点 方 法

1. 取整数 $a_i(1\leqslant i\leqslant s)$ 满足 $a_1=1, 1<a_i<q(2\leqslant i\leqslant s), a_i\neq a_j(i\neq j)$ 及 $g$、$c$、$d$、$(a_i,q)=1(1\leqslant i\leqslant s)$ 则布点形式为

$$P_n(k)=(ka_1,ka_2,\cdots,ka_s)(\bmod q),\quad k=1,2,\cdots,q. \tag{2.1}$$

显然 $s\leqslant\varphi(q)$, 此处 $\varphi(q)$ 为 Euler 函数, 即不大于 $q$ 且与 $q$ 互素的自然数的个数, 这 $\varphi(q)$ 个数称为模 $q$ 的缩系. 所以如果确定水平数为 $q$, 则 $\varphi(q)$ 表示用 (2.1) 式的点列安排试验的最多因素个数. 我们用 $\boldsymbol{a}=(a_1,\cdots,a_s)$ 表示对应于 (2.1) 式的整向量.

**引理 2.1** 假定 $q=p_1^{l_1}\cdots p_m^{l_m}$, 此处 $p_1<\cdots<p_m$ 为素数, 则

$$\varphi(q)=q\prod_{i=1}^{m}\left(1-\frac{1}{p_i}\right).$$

今后我们用 $p$ 表示素数, 由引理 2.1 可见 $\varphi(q)\leqslant q-1$, 且只有当 $q=p$ 为素数时取等号. 选择 $a_1,\cdots,a_s$ 的原则是下节所述的一致分布理论. 但当 $\varphi(q)$ 较大时,

* 原载《科学通报》第 2 期, 1981 年, 65~70.

从缩系中挑选 $a_1,\cdots,a_s$, 共有 $\binom{\varphi(q)-1}{s-1}$ 种可能性 (因为 $a_1=1$), 计算量仍很大.

2. 当 $q=p$ 时, 建议用下面的点

$$Q_p(k)=(k,kb,\cdots,kb^{s-1})(\operatorname{mod}p),\quad k=1,2,\cdots,p, \tag{2.2}$$

此处 $b$ 为适合于 $1<b<p$ 且 $b^i\not\equiv b^j(\operatorname{mod}p)(i\neq j)$ 的整数, 所以我们常常取 $b$ 为原根, 我们用 $\boldsymbol{b}=(1,b,\cdots,b^{s-1})$ 表示对应于 (2.2) 式的整向量.

3. 当 $q=p-1$ 时, 用 (2.2) 式去掉最后一个坐标, 往往比直接用模 $q$ 的缩系, 由 (2.1) 式形式的点造表更好.

## §3. 一致分布

命 $G_s$ 表示 $s$ 维单位立方体, 命 $1<n_1<n_2<\cdots$ 表示任意一个自然数列及

$$R_{n_l}(k)=(x_1^{(n_l)}(k),\cdots,x_s^{(n_l)}(k))(1\leqslant k\leqslant n_l)$$

表示 $G_s$ 中的点列. 对于任意 $\boldsymbol{r}=(r_1,\cdots,r_s)\in G_s$, 命 $N_{n_l}(\boldsymbol{r})$ 表示适合于

$$0\leqslant x_i^{(n_l)}(k)<r_i\quad(1\leqslant i\leqslant s)$$

的点 $R_{n_l}(k)(1\leqslant k\leqslant n_l)$ 的个数. 如果当 $n_l\to\infty$ 时有

$$\sup_{r\in G_s}\left|\frac{N_{n_l}(\boldsymbol{r})}{n_l}-|\boldsymbol{r}|\right|=o(1),$$

此处 $|\boldsymbol{r}|=\prod_{i=1}^{s}r_i$, 则称点列序列 $\{R_{n_l}(k)\}(1<n_1<n_2<\cdots)$ 在 $G_s$ 上一致分布.

略去指标 $l$, 命

$$\sup_{r\in G_s}\left|\frac{N_n(\boldsymbol{r})}{n}-|\boldsymbol{r}|\right|=D(n).$$

$D(n)$ 称为 $R_n(k)(1\leqslant k\leqslant n)$ 的偏差, 则我们要选择 $\boldsymbol{a}$ 及 $\boldsymbol{b}$ 使点列

$$\left(\left\{\frac{a_1k}{q},\cdots,\frac{a_sk}{q}\right\}\right)\quad(1\leqslant k\leqslant q) \tag{3.1}$$

与

$$\left(\left\{\frac{k}{p}\right\},\left\{\frac{kb}{p}\right\},\cdots,\left\{\frac{kb^{s-1}}{p}\right\}\right)\quad(1\leqslant k\leqslant p) \tag{3.2}$$

所对应的偏差尽可能地小. 我们用 $D(q,\boldsymbol{a})$ 和 $D(p,\boldsymbol{b})$ 分别表示 (3.1) 和 (3.2) 式的偏差. 形如 (3.1) 式的点首先是 коробов[2] 与 Hlawka[3] 引进的, 形如 (3.2) 式的点则为 коробов[4] 所引进.

命 $\bar{x}=\max(1,|x|)$, $\|\boldsymbol{x}\|=\bar{x}_1\cdots\bar{x}_s$ 及 $(\boldsymbol{x},\boldsymbol{y})=\sum_{i=1}^{s}x_iy_i$ 表示向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的内积.

**引理 3.1** (Erdös-Turan-Koksma). 设 $\eta>0$, 整数 $h\geqslant 2$, 则

$$D(n)=\sum_{|m_i|\leqslant h}{}' \frac{1}{\|\pi\boldsymbol{m}\|}\left|\frac{1}{n}\sum_{k=1}^{n}e^{2\pi i(\boldsymbol{m},R_n(k))}\right|+O(\eta)+O(\eta^{-1}(\ln h)^{s-1}),$$

此处 $\Sigma'$ 表示求和号中去掉 $m_1=\cdots=m_s=0$ 一项 (证明见文献 [5]).

**引理 3.2** 假定 $(a_i,q)=1(1\leqslant i\leqslant s)$, 则对于任何整数 $r\geqslant 1$ 皆有

$$\sum_{(\boldsymbol{a},\boldsymbol{m})\equiv 0(\operatorname{mod} q)}{}' \frac{1}{\|\pi\boldsymbol{m}\|}-\sum_{\substack{(\boldsymbol{a},\boldsymbol{m}^{(0)})\equiv 0(\operatorname{mod} q)\\ -q/2>\boldsymbol{m}_i^{(0)}<q/2}}{}' \frac{1}{\|\pi\boldsymbol{m}^{(0)}\|}=O(q^{-1}(\ln rq)^{s-1})$$

(证明见文献 [5]).

由引理 3.1 与引理 3.2 可得

**引理 3.3** $D(q,\boldsymbol{a})=\sum_{\substack{(\boldsymbol{a},\boldsymbol{m})\equiv 0(\operatorname{mod} q)\\ |m_i|\leqslant q/2}}{}' \dfrac{1}{(\pi m)}+O(q^{-1}(\ln q)^{s})$.

**引理 3.4** 设 $x\in(0,1)$, 则

$$\sum_{|m|<q}e^{2\pi imx}/\overline{\pi m}=1-\frac{2}{\pi}\ln(2\sin\pi x)+\frac{\vartheta}{\pi q\langle x\rangle},$$

此处 $|\vartheta|\leqslant 1$ 及 $\langle x\rangle=\min(\{x\},1-\{x\})$, 其中 $\{x\}$ 表示 $x$ 的分数部分.

**定理 3.1**

$$D(q,\boldsymbol{a})=\frac{1}{q}\sum_{k=1}^{q-1}\prod_{\nu=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\left\{\frac{a_\nu k}{q}\right\}\right)\right)-1+O(q^{-1}(\ln q)^{s}).\tag{3.3}$$

**证** 由引理 3.1, 3.3 与 3.4 可得

$$\begin{aligned}D(q,\boldsymbol{a})=&\frac{1}{q}\sum_{k=1}^{q}\sum_{|m_i|\leqslant q/2}\frac{e^{2\pi i(\boldsymbol{a},\boldsymbol{m})k/q}}{\|\pi\boldsymbol{m}\|}-1+O(q^{-1}(\ln q)^{s})\\ =&\frac{1}{q}\sum_{k=1}^{q-1}\prod_{\nu=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\left\{\frac{a_\nu k}{q}\right\}\right)\right)-1-\frac{1}{q}\sum_{|m_i|\leqslant q/2}\frac{1}{\|\pi\boldsymbol{m}\|}\\ &+O(q^{-1}(\ln q)^{s})+O\left(q^{-1}(\ln q)^{s-1}\sum_{k=0}^{q-1}\sum_{\nu=1}^{s}\frac{1}{q\left\langle\dfrac{a_\nu k}{q}\right\rangle}\right)\end{aligned}$$

$$=\frac{1}{q}\sum_{k=1}^{q-1}\prod_{\nu=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\left\{\frac{a_\nu k}{q}\right\}\right)\right)-1+O(q^{-1}(\ln q)^s).$$

定理证完.

在 (3.3) 式右端的第二项与第三项分别是 $-1$ 与 $O(q^{-1}(\ln q)^s)=o(1)$ (当 $q\to\infty$). 所以比较诸 $D(q,\boldsymbol{a})$ 的大小即归结为比较

$$D_1(q,\boldsymbol{a})=\frac{1}{q}\sum_{k=1}^{q-1}\prod_{\nu=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\left\{\frac{a_\nu k}{q}\right\}\right)\right)$$

的大小, 但当 $q$ 较小时, 诸 $D_1(q,\boldsymbol{a})$ 不易分辨, 所以我们引进

$$D_2(q,\boldsymbol{a})=\frac{1}{q}\sum_{k=1}^{q}\prod_{\nu=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\nu k}}{(q+1)}\right)\right),$$

此处 $a_{\nu k}\equiv a_\nu k(\operatorname{mod} q), 1\leqslant a_{\nu k}<q(1\leqslant k\leqslant q)$, 并规定 $a_\nu q\equiv q(\operatorname{mod} q)$.

**定理 3.2**　$D_2(q,\boldsymbol{a})=D_1(q,\boldsymbol{a})+O(q^{-1}(\ln q)^s)$.

**证**　显然

$$D_2(q,\boldsymbol{a})=J+O(q^{-1}(\ln q)^s),$$

此处

$$J=\frac{1}{q}\sum_{k=1}^{q-1}\prod_{\nu=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\nu k}}{q+1}\right)\right),$$

由于

$$\begin{aligned}
&\prod_{v=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\nu k}}{q}\right)\right)-\prod_{\nu=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\nu k}}{q+1}\right)\right)\\
=&\sum_{r=0}^{s-1}\left[\prod_{\mu=1}^{r}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\mu k}}{q+1}\right)\right)\prod_{\nu=r+1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\nu k}}{q}\right)\right)\right.\\
&\left.-\prod_{\mu=1}^{r+1}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\mu k}}{q+1}\right)\right)\prod_{\nu=r+2}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\nu k}}{q}\right)\right)\right]\\
=&\sum_{r=0}^{s-1}\prod_{\mu=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\nu k}}{q+1}\right)\right)\prod_{\nu=r+2}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\nu k}}{q}\right)\right)\\
&\times\left(\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{r+1,k}}{q+1}\right)-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{r+1,k}}{q}\right)\right)
\end{aligned}$$

及

$$\left|\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{r+1,k}}{q+1}\right)-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{r+1,k}}{q}\right)\right|$$

$$=\frac{2}{\pi}\left|\int_{a_{r+1,k/(q+1)}}^{a_{r+1,k/q}} d\ln\sin\pi x\right| = O\left(\frac{a_{r+1,k}}{q^2\left\langle\frac{a_{r+1,k}}{q}\right\rangle}\right),$$

所以

$$\prod_{\nu=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\nu k}}{q}\right)\right)-\prod_{\nu=1}^{s}\left(1-\frac{2}{\pi}\ln\left(2\sin\pi\frac{a_{\nu k}}{q+1}\right)\right)$$

$$=O\left(\sum_{r=0}^{s-1}\frac{a_{r+1,k}}{q^2\left\langle\frac{q_{r+1,k}}{q}\right\rangle}(\ln q)^{s-1}\right)$$

从而

$$|D_2(q,\boldsymbol{a})-J| = O\left(q^{-1}\sum_{k=1}^{q-1}\sum_{r=1}^{s}\frac{a_{rk}}{q^2\left\langle\frac{a_{rk}}{q}\right\rangle}(\ln q)^{s-1}\right)$$

$$= O\left(q^{-1}\sum_{l=1}^{q}(\ln q)^{s-1}l^{-1}\right) = O(q^{-1}(\ln q)^s).$$

定理证完.

使 $D_2(q,\boldsymbol{a})$ 及 $D_2(p,\boldsymbol{b})$ 达到极小的整数向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$, 即可用来构造点列 (3.1) 与 (3.2) 式, 即点列 (2.1) 与 (2.2) 式, 用它们安排试验的方法称为均匀设计

## §4. 均匀设计表

由于同余式 $(\boldsymbol{a},\boldsymbol{m})\equiv 0(\mathrm{mod}\, q)$ 可以写成 $a_\mu^{-1}(\boldsymbol{a},\boldsymbol{m})\equiv 0(\mathrm{mod}\, q)$ 及 $(\boldsymbol{b},\boldsymbol{m})\equiv 0(\mathrm{mod}\, p)$ 可以写成 $b^{-s+1}(\boldsymbol{b},\boldsymbol{m})\equiv 0(\mathrm{mod}\, p)$, 所以在比较 $D_2(q,\boldsymbol{a})$ 或 $D_2(p,\boldsymbol{b})$ 时, 工作量还可以减少.

1. 当 $q=p$ 采用布点 (2.2) 式, 去掉最后一点, 即得到对应于 $q-1$ 的均匀设计表, 所以表 4.1 实际上给出了 $q=4,5,6,7,10,11,12,13,16,17,18,19,22,23,28,29,30,31$ 的均匀设计表.

2. 适合于 $4\leqslant q\leqslant 31$ 的正整数, 除上述外, 还有 $q=8,9,14,15,20,21,24,25,26,27$, 其中 $q$ 为偶数的均匀设计表, 可以由奇数 $q+1$ 的均匀设计表去掉最后一个试验点来得到, 故只要给出 $q=9,15,21,25,27$ 对应的均匀设计表 (表 1). 当 $q=9,15,21$ 时, 采用点列 (2.1) 式, 当 $q=25,27$ 时, 由于 $\mathrm{mod}\, q$ 有原根, 故采用点列 (2.2) 式, 结果见表 2.

3. 例: 取 $s=4, q=5$, 则由表 1 可得点列

$$(k,2k,2^2k,2^3k)(\mathrm{mod}\, 5),\quad 1\leqslant k\leqslant 5,$$

或列成表 3 的形式, 其记号 $U_5(5^4)$ 是类似于正交试验法中的 $L_{25}(5^6)$.

均匀设计的数据分析, 常借助于回归分析或逐步回归.

**表 1　素数 $p$ 及 $p-1$ 的均匀设计表**

| $p$ \ $s$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 2 | 2 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 7 | 3 | 3 | 3 | 3 | 3 | | | | | | | | | | | | | | | | | | | | | | | | |
| 11 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | | | | | | | | | | | | | | | | | | | | |
| 13 | 5 | 4 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | | | | | | | | | | | | | | | | | | |
| 17 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | | | | | | | | | | | | | |
| 19 | 8 | 8 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | | | | | | | | | | | | |
| 23 | 7 | 17 | 17 | 17 | 17 | 17 | 15 | 15 | 15 | 15 | 15 | 7 | 7 | 17 | 17 | 17 | 17 | 17 | 7 | 7 | 7 | | | | | | | | |
| 29 | 12 | 9 | 16 | 16 | 16 | 16 | 8 | 8 | 8 | 8 | 8 | 14 | 14 | 14 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 18 | 18 | 18 | 18 | | |
| 31 | 12 | 22 | 22 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |

**表 2　$q$ 和 $q-1$ 的均匀设计表**

| $q$ \ $s$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 4 | 4,7 | 4,7,2 | 4, 7,<br>2, 5 | 4, 7, 2,<br>5, 8 | | | | | | | | | | | | | | |
| 15 | 11 | 4,7 | 4,7,13 | 4, 7,<br>13, 2 | 2, 4,<br>7, 11,<br>14 | 2, 4,<br>7, 11,<br>14, 13 | 2, 4, 7,<br>11, 14,<br>13, 8 | | | | | | | | | | | | |
| 21 | 13 | 4,10 | 4,10,13 | 1, 10,<br>16, 19 | 4, 10,<br>13, 16,<br>19 | 4, 10,<br>13, 16,<br>19, 20 | 4, 5, 8,<br>10, 11,<br>17, 19 | 2, 4, 5,<br>8, 10, 11,<br>17, 19 | 2, 4, 5,<br>8, 10,<br>11, 16,<br>17, 19 | 2, 4, 5,<br>8, 10,<br>11, 13,<br>16, 17,<br>19, | 2, 4, 5,<br>8, 10,<br>11, 13,<br>16, 17,<br>19, 20 | | | | | | | | |
| 25 | 11 | 11 | 11 | 11 | 4 | 9 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 27 | 8 | 8 | 20 | 20 | 20 | 16 | 16 | 16 | 20 | 5 | 5 | 20 | 20 | 20 | 5 | 5 | 5 | | |

**表 3　$U_5(5^4)$**

| No \ 因子 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1 | 2 | 4 | 3 |
| 2 | 2 | 4 | 3 | 1 |
| 3 | 3 | 1 | 2 | 4 |
| 4 | 4 | 3 | 1 | 2 |
| 5 | 5 | 5 | 5 | 5 |

## 参 考 文 献

[1] 中国科学院数学研究所统计组, 方差分析, 科学出版社, 1977.
[2] Коробов Н. М., ДАНСССР, **124**, 6(1959), 1207~1210.
[3] Hlawka, E., *Mon. Math.*, **66**, 2(1962), 140~151.
[4] Коробов Н. М., *Вес. МГу*, 4, (1959), 19~25.
[5] 华罗庚, 王元, 数论在近似分析中的应用, 科学出版社, 1978.

# 应用统计中的数论方法 (I)*

王元 方开泰
(中国科学院数学研究所) (中国科学院应用数学研究所)

(纪念数学年刊创刊十周年)

## 提　　要

本文将给出数论方法 (伪蒙特卡罗方法) 在某些特殊区域, 例如立方体, 球面, 球, 单纯形等上的连续多元分布的概率与矩的数值计算问题上的应用, 在此我们将给出这些区域中的均匀分布点集. 这一方法还可以应用于试验设计、模拟、几何概率与最优化.

## §1. 导　　言

概率与矩的数值计算问题实际上就是数值积分问题. 多重积分近似计算与最优化的数论方法 (或伪蒙特卡罗方法) 基于一致分布 (或均匀分布)$(u,d)$ 理论. 命 $K=[a_1\times b_1]\times\cdots\times[a_s,b_s]$ 为 $R^s$ 中的一个立方体, $\boldsymbol{b}=(b_1,\cdots,b_s)'$, $\boldsymbol{x}=(x_1,\cdots,x_s)'$ 及 $F(\boldsymbol{x})$ 为 $K$ 上一个连续单调分布函数且满足 $F(\boldsymbol{b})=1$ 及当至少有一个 $x_i=a_i$ 时, $F(\boldsymbol{x})=0$. 注意 $a_i's$ 与 $b_i's$ 可以分别取 $-\infty$ 与 $\infty$, 我们用 $\boldsymbol{x}\leqslant\boldsymbol{b}$ 表示 $x_i\leqslant b_i(i=1,\cdots,s)$. 对于 $K$ 中一个点集 $P=(\boldsymbol{x}_k,k=1,\cdots,n)$ 及一个立方体 $G=[a_1,x_1]\times\cdots\times[a_s,x_s]$, 此处 $\boldsymbol{x}\leqslant\boldsymbol{b}$, 命 $N(P,G)$ 表示 $P$ 中适合 $\boldsymbol{x}_k\in G$ 的点数, 命

$$\sup_G\left|\frac{N(P,G)}{n}-F(\boldsymbol{x})\right|=D_F(n,P).$$

$D_F(n,P)$ 称为 $P$ 关于 $F(\boldsymbol{x})$ 的 $F$- 偏差. 若 $P_n=(x_1^{(n)},\cdots,x_{k_n}^{(n)})$ 为 $K$ 中一个贯, 其中当 $n\to\infty$ 时 $k_n\to\infty$, 及当 $n\to\infty$ 时 $D_F(k_n,P)=o(1)$, 则称 $P_n$ 为 $F$- 一致分布贯. 若 $K=I^s$, 此处 $I=[0,1]$, 及 $F(x)$ 为 $I^s$ 上的均匀分布, 即 $F(x)=x_1\cdots x_s$, 则在上面记号中可以略去 $F$(见 Weyl[12], Hlawka 与 Much [5] 及 Niederreiter [9]).

---

* 原载 "数学年刊", 11B, 1, 1990, 51～65.

这一工作受到中国国家自然科学基金及中国科学院的基金支持.

若 $P=\{\boldsymbol{x}_k,k=1,\cdots,n\}$ 为 $I^s$ 的一个集合及有偏差 $D(n,P)$ 或简单记为 $D(n)$, 又若 $f(\boldsymbol{x})$ 为在 Hardy 与 Krause 意义下的有界变差函数, 且有全变差 $V(f)$, 则

$$\left|\int_{I^s} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}-\frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{x}_k)\right| \leqslant V(f)D(n). \tag{1.1}$$

(见 Koksma [7], Hlawka [4], 华罗庚与王元 [6].)

命 $D$ 为 $R^s$ 中的一个区域 (例如球、球面、单纯形等), 在本文中, 我们将更多地阐述积分

$$I=\int_D f(\boldsymbol{z})\mathrm{d}v \tag{1.2}$$

的数值计算, 此处 $\mathrm{d}v$ 为 $D$ 的体积元素及 $D$ 有一个参数表示. 首先我们可以用一个变换将 $D$ 上的求积公式变成 $I^t$ 上的求积公式. 此处 $t$ 为 $D$ 的维数. 另一个处理这一问题的方法为利用 $D$ 上的一致分布贯. 我们将从 $I^t$ 的一个一致分布贯出发, 然后导出关于某些分布函数的 u. d 贯, 特别是关于某些特殊区域 (球, 球面, 单纯形等) 的均匀分布函数的 u. d 贯, 这些贯在模拟, 几何概率, 试验设计及统计中的许多问题中都经常被用到, 更多的细节将在同样标题的下一篇文章中讲述.

$D$ 上 u. d 贯的另一个应用为最优化, 命 $f(x)$ 为 $D$ 上的一个连续函数, 我们欲求出它在 $D$ 上的全局极大值 $M$. 关于这个极值问题有许多梯度法可以处理 (见 Avriel [2]). 但不幸地是只有很少情况下才能得到全局极大, 而当 $f$ 非单峰及 $D$ 的维数较大, 例如 $\geqslant 5$ 时, 一般只能得到一个局部极大. 这是由于一般说来, 解答依赖于初始点的选取. 因此我们用下面的程序来寻求 $M$ 的近似值:

$$m_1=f(\boldsymbol{x}_1),$$

$$m_{k+1}=\begin{cases} m_k, & \text{当 } f(\boldsymbol{x}_{k+1})\leqslant m_k,\\ f(\boldsymbol{x}_{k+1}), & \text{当 } f(\boldsymbol{x}_{k+1})\geqslant m_k,\end{cases}$$

此处 $(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots)$ 为 $D$ 上的一个 u. d 贯. 即 $P_n=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$ 为一个 u. d 贯. 经过 $n$ 步之后, 若 $f(\boldsymbol{x})$ 适合某些正则条件, 则可以相信 $m_n$ 接近于 $M$. 我们经常用量

$$d(n,D)=\max_{x\in D}\min_{1\leqslant k\leqslant n} d(\boldsymbol{x},\boldsymbol{x}_k), \tag{1.3}$$

来度量 $P_n$ 的均匀性, 此处 $d(\boldsymbol{x},\boldsymbol{x}_k)$ 表示 $\boldsymbol{x}$ 与 $\boldsymbol{x}_k$ 间的欧氏矩离. $d(n,D)$ 称为集合 $\{\boldsymbol{x}_k,k=1,\cdots,n\}$ 的离差. 若 $D=I^s$, 则可以证明

$$\sqrt{s}n^{-1/s}\leqslant d(n,D)\leqslant 2\sqrt{s}(D(n))^{1/s}. \tag{1.4}$$

(见 Zielinski[13] 及 Niederreiter[10].) 这表示当 $n$ 大时, $m_n$ 的确接近于 $M$. 在第 4 节中, 我们将上述结果推广至某些 $D$ 并给出在统计中的应用.

## §2. 数值积分

命 $D$ 表示 $R^s$ 上的一个有界区域, 我们要求计算积分 (1.2), 假定 $D$ 的维数为 $s$, $\mathrm{d}v = \prod_{i=1}^{s} \mathrm{d}x_i = \mathrm{d}\boldsymbol{x}$ 及 $D \subset T^s$, 则可以简单地介绍下面的公式

$$I = \int_{I^s} f(\boldsymbol{x}) I_D(\boldsymbol{x}) \mathrm{d}\boldsymbol{x},$$

此处 $I_D(x)$ 为 $D$ 的指标函数 (见华罗庚与王元 [6]), 由于 $f(\boldsymbol{x})I_D(\boldsymbol{x})$ 可能在 $D$ 的边界上不连续, 所以这一方法有时会导致很大的误差. 在统计中, $D$ 常常是很特殊的, 所以可以将 $D$ 上的积分转变成 $I^t(t \leqslant s)$ 上的积分. 进而言之, 假定 $D$ 有一个表示

$$x_j = x_j(\varphi_1, \cdots, \varphi_t) = x_j(\boldsymbol{\varphi}), j = 1, \cdots, s, \tag{2.1}$$

此处 $\varphi \in I^t$ 及 $x_j$ 关于 $\varphi_i$ 在 $I$ 上有连续导数 $(j = 1, \cdots, s, i = 1, \cdots, t)$. 命

$$\boldsymbol{T} = (\partial x_j / \partial \varphi_i), i = 1, \cdots, t, j = 1, \cdots, s,$$

及

$$J(\boldsymbol{\varphi}) = \det(\boldsymbol{T}\boldsymbol{T}')^{1/2}.$$

当 $t = s$ 时, $J(\varphi)$ 就是 $\boldsymbol{x}$ 关于 $\boldsymbol{\varphi}$ 的 Jacobian. 则

$$I = \int_D f(\boldsymbol{x}) \mathrm{d}v = \int_{I^t} f(\boldsymbol{x}(\boldsymbol{\varphi})) J(\boldsymbol{\varphi}) \mathrm{d}\boldsymbol{\varphi}, \tag{2.2}$$

此处 $\mathrm{d}\varphi = \prod_{i=1}^{t} \mathrm{d}\varphi_i$. 因此 $I^t$ 上的一个求积公式即诱导出 $D$ 上的一个求积公式, 记 $v(D)$ 为 $D$ 的体积, 则

$$v(D) = \int_{I^t} J(\boldsymbol{\varphi}) \mathrm{d}(\boldsymbol{\varphi}). \tag{2.3}$$

假定 $\varphi_1, \cdots, \varphi_t$ 是独立的及

$$v(D)^{-1} J(\boldsymbol{\varphi}) = \prod_{i=1}^{t} f_i(\varphi_i),$$

此处 $f_i(\varphi_i)$ 为 $\varphi_i$ 的密度函数. $i = 1, \cdots, t$ 及其对应的分布函数为

$$F_i(x_i) = \int_0^{x_i} f_i(\varphi_i) \mathrm{d}\varphi_i, i = 1, \cdots, t,$$

满足 $F_i(0)=0$ 及 $F_i(1)=1, i=1,\cdots,t$. 命 $F_i(x_i)=y_i$ 及 $F_i^{-1}(y_i)$ 表示 $F(x_i)$ 的逆函数, $i=1,\cdots,t$. 则

$$\int_{I^t} f(\boldsymbol{\varphi})J(\boldsymbol{\varphi})\mathrm{d}\boldsymbol{\varphi} = v(D)\int_{I^t} f(\boldsymbol{x}(\boldsymbol{F}^{-1}(\boldsymbol{y}))\mathrm{d}\boldsymbol{y}, \tag{2.4}$$

此处 $F^{-1}(\boldsymbol{y})=(F_1^{-1}(y_1),\cdots,F_t^{-1}(y_t))^{-1}$ 及 $\mathrm{d}\boldsymbol{y}=\prod_{i=1}^{t}\mathrm{d}y$.

对于给定 $I^t$ 上有偏差 $D(n)$ 的一个集合 $\{\boldsymbol{b}_k=(b_{k_1},\cdots,b_{k_t})', k=1,\cdots,n\}$, 我们得到具有 $F$- 偏差 $D_F(\boldsymbol{n},\{\boldsymbol{c}_k\})=D(n)$ 的一个集合 $\{\boldsymbol{C}_k=F^{-1}(\boldsymbol{b}_k), k=1,\cdots,n\}$, 其中 $F(\boldsymbol{x})=\prod_{i=1}^{t}F_i(x_i)$. 由 (1.1), (2.2) 与 (2.4) 得

$$\begin{aligned}&\left|\int_D f(\boldsymbol{x})\mathrm{d}v - v(D)\frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{x}(\boldsymbol{F}^{-1}(\boldsymbol{b}_k)))\right| \\ &\leqslant v(D)D_F(n,\{\boldsymbol{c}_k\})V(f(\boldsymbol{F}^{-1})) \\ &= v(D)D(n)V(f).\end{aligned} \tag{2.5}$$

由 (2.2), (2.5) 及数论方法, 我们可以得到下面关于重积分 (1.2) 的两个近似计算公式

$$\int_D f(\boldsymbol{x})\mathrm{d}v \cong \frac{1}{n}\sum_{k=1}^{n} f(\boldsymbol{x}(\boldsymbol{b}_k))J(\boldsymbol{b}_k) \tag{2.6}$$

与

$$\begin{aligned}\int_D f(\boldsymbol{x})\mathrm{d}v &\cong \frac{1}{n}v(D)\sum_{k=1}^{n} f(\boldsymbol{x}(\boldsymbol{c}_k)) \\ &= \frac{1}{n}v(d)\sum_{k=1}^{n} f(\boldsymbol{x}(\boldsymbol{F}^{-1}(\boldsymbol{b}_k))).\end{aligned} \tag{2.7}$$

这两个公式的精密度有相同的阶.

定义 $D$ 上的一个集合

$$P=\{\boldsymbol{x}_k=\boldsymbol{x}(\boldsymbol{c}_k), k=1,\cdots,n\}. \tag{2.8}$$

区域 $D$ 的由 $\boldsymbol{\varphi}\leqslant\boldsymbol{y}$ 定义的部分的体积 $v(\boldsymbol{\varphi}\leqslant\boldsymbol{y})$ 为

$$v(\boldsymbol{\varphi}\leqslant\boldsymbol{y}) = \int_{\varphi\leqslant y} J(\boldsymbol{\varphi})\mathrm{d}\boldsymbol{\varphi} = v((D)\prod_{i=1}^{t}F_i(y_i),$$

所以

$$v(\boldsymbol{v}\leqslant\boldsymbol{y})/v(D)=\prod_{i=1}^{t}F_i(y_i).$$

因此, 若要求集合 $P$ 在 $D$ 上一致散布, 或 $P$ 中位于由 $\boldsymbol{\varphi}\leqslant\boldsymbol{y}$ 定义的区域中的点数 $N(\boldsymbol{\varphi}\leqslant\boldsymbol{y})$ 与 $n$ 的比渐近地等于 $v(\boldsymbol{\varphi}\leqslant\boldsymbol{y})$ 与 $v(D)$ 的比, 我们需取集合 $\{\boldsymbol{c}_k,k=1,\cdots,n\}$ 在 $I^t$ 上有较低的 $F$- 偏差. 由于 $\{\boldsymbol{c}_k,k=1,\cdots,n\}$ 有 $F$- 偏差 $D(n)$, 所以

$$\begin{aligned}\sup_{\boldsymbol{y}\in I^t}\left|\frac{N(\boldsymbol{\varphi}\leqslant\boldsymbol{y})}{n}-\frac{v(\boldsymbol{\varphi}\leqslant\boldsymbol{y})}{v(D)}\right|&=\sup_{\boldsymbol{y}\in I^t}\left|\frac{N(\boldsymbol{\varphi}<\boldsymbol{y})}{n}-F(\boldsymbol{y})\right|\\&=D(n).\end{aligned}\tag{2.9}$$

这样我们就建议了一个由 $I^t$ 上有低偏差的集合导出 $D$ 上一致散布的集合 $P$ 的程序, 现在我们来举一些例子:

**例 1** 假定 $D$ 为单纯形 $A_s=\{\boldsymbol{x}:0\leqslant x_s\leqslant x_{s-1}\leqslant\cdots\leqslant x_1\leqslant 1\}$. 则 $D$ 有表示

$$x_j=\varphi_1\cdots\varphi_j,j=1,\cdots,s,$$

此处 $\boldsymbol{\varphi}\in I^t$, 所以

$$J(\boldsymbol{\varphi})=\prod_{i=1}^{s-1}\varphi_i^{s-i},$$

及

$$v(A_s)=\int_{I^s}J(\boldsymbol{\varphi})\mathrm{d}\boldsymbol{\varphi}=\prod_{i=1}^{s-1}\frac{1}{s-i+1}=1/s!.$$

因此

$$f_i(\varphi_i)=(s-i+1)\varphi_i^{s-i},i=1,\cdots,s$$

为对应于分布函数

$$F_i(x_i)=\int_0^{x_i}f_i(\varphi_i)\mathrm{d}\varphi_i=x_i^{s-i+1}$$

的密度函数. 对于 $I^s$ 上一个有偏差 $D(n)$ 的集合 $\{b_k,k=1,\cdots,n\}$, 我们有一个 $I^s$ 上有 $F$- 偏差 $D(n)$ 的集合

$$\boldsymbol{c}_k=\boldsymbol{F}^{-1}(\boldsymbol{b}_k)=(b_{k1}^{1/s},b_{k2}^{1/(s-1)},\cdots,b_{k,s-1}^{1/2},b_{ks}),k=1,\cdots,n.$$

从而有一个 $A_s$ 的集合 $P$:

$$\boldsymbol{x}_k=(x_{k1},\cdots,x_{ks})',k=1,\cdots,n,\tag{2.10}$$

此处

$$x_{kj}=\prod_{i=1}^{j} b_{ki}^{1/(s-i+1)}, k=1,\cdots,n, j=1,\cdots,s. \tag{2.11}$$

集合 $P$ 适合 (2.9).

**例 2** 命 $D$ 为 $s$- 维单位球

$$B_s=\{\boldsymbol{x}: x_1^2+\cdots+x_s^2 \leqslant 1\}$$

它有表示

$$\begin{aligned}x_j&=\varphi_1 S_2\cdots S_j C_{j+1}, j=1,\cdots,s-1\\x_s&=\varphi_1 S_2\cdots S_{s-1} S_s,\end{aligned}$$

此处 $S_k=\sin(\pi\varphi_k)$, $C_k=\cos(\pi\varphi_k)$, $k=2,\cdots,s-1$, $S_s=\sin(2\pi\varphi_s)$ 及 $C_s=\cos(2\pi\varphi_s)$, 其中 $\boldsymbol{\varphi}\in I^s$. 于是得

$$J(\boldsymbol{\varphi})=2\pi^{s-1}\varphi_1^{s-1}\prod_{i=2}^{s} S_i^{s-i}$$

及

$$\begin{aligned}v(B_s)&=\int_{I^s} J(\boldsymbol{\varphi})\mathrm{d}\boldsymbol{\varphi}\\&=2\pi^{s-1}\int_0^1 \varphi_1^{s-1}\mathrm{d}\varphi_1\prod_{i=2}^{s}\int_0^1 S_i^{s-i}\mathrm{d}\varphi_i\\&=\frac{2}{s}\prod_{i=2}^{s} B\left(\frac{1}{2},\frac{s-i+1}{2}\right),\end{aligned}$$

其中用到对于任何整数 $m>0$ 有

$$\int_0^1(\sin(\pi x))^m\mathrm{d}x=\frac{1}{\pi}B\left(\frac{1}{2},\frac{m+1}{2}\right),$$

因此

$$f_i(\varphi_i)=\begin{cases} s\varphi_1^{s-1}, & \text{当 } i=1,\\ \pi(\sin(\pi\varphi_i))^{s-i}\Big/B\left(\dfrac{1}{2},\dfrac{s-i+1}{2}\right), & \text{当 } i=2,\cdots,s\end{cases}$$

为 $I^s$ 上密度函数, 其对应的分布函数为

$$F_1(x_1)=s\int_0^{x_1}\varphi_i^{s-1}\mathrm{d}\varphi_1=x_1^s,$$

$$F_i(x_i) = \frac{\pi}{B(1/2, (s-i+1))/2} \int_0^{x_i} (\sin \pi x)^{s-i} \mathrm{d}x, i = 2, \cdots, s.$$

对于 $I^s$ 上的一个有偏差 $D(n)$ 的集合 $\{\boldsymbol{b}_k, k = 1, \cdots, n\}$, 我们 $I^s$ 上有 $F$- 偏差 $D(n)$ 的集合 $\{\boldsymbol{c}_k, k = 1, \cdots, n\}$, 此处

$$c_{k1} = b_{k1}^{1/s},$$
$$F_i(c_{ki}) = b_{ki}, i = 2, \cdots, s, k = 1, \cdots, n,$$

最后得到 $B_s$ 上的一个集合 $P$:

$$\boldsymbol{x}_k = (x_{k1}, \cdots, x_{ks})', k = 1, \cdots, n, \tag{2.12}$$

此处

$$x_{kj} = b_{k1}^{1/s} \prod_{i=2}^{j} S_{ki} C_{k,j+1}, j = 1, \cdots, s-1,$$
$$x_{ks} = b_{k1}^{1/s} \prod_{i=2}^{s} S_{ki}, \tag{2.13}$$

其中 $S_{ki} = \sin(\pi c_{ki}), C_{ki} = \cos(\pi c_{ki})$, $i = 1, \cdots, s-1$, $S_{ks} = \sin(2\pi c_{ks})$, $C_{ks} = \cos(2\pi c_{ks})$, $k = 1, \cdots, n$.

**例 3** 命 $D$ 为 $s-1$ 维单位球面

$$S^{s-1} = \{\boldsymbol{x} : x_1^2 + \cdots + x_s^2 = 1\},$$

它有表示

$$x_j = \prod_{i=1}^{j-1} S_i C_j, j = 1, \cdots, s-1,$$
$$x_s = \prod_{i=1}^{s-1} S_i,$$

此处 $S_i = \sin(\pi\varphi_i)$, $C_i = \cos(\pi\varphi_i)$, $i = 1, \cdots, s-2$, $S_{s-1} = \sin(2\pi\varphi_{s-1})$ 及 $C_{s-1} = \cos(2\pi\varphi_{s-1})$, 其中 $\boldsymbol{\varphi} \in I^{s-1}$. 则

$$J(\boldsymbol{\varphi}) = 2\pi^{s-1} \prod_{i=1}^{s-2} S_i^{s-i-1}$$

及

$$v(S^{s-1}) = \int_{I^{s-1}} J(\boldsymbol{\varphi}) \mathrm{d}\boldsymbol{\varphi} = 2\pi \prod_{i=1}^{s-2} \left(\frac{1}{2}, \frac{s-i}{2}\right).$$

所以

$$f_i(\varphi_i) = \pi S_i^{s-i-1}/B(1/2,(s-i)/2), i=1,\cdots,s-1$$

为 $I$ 上的密度函数, 其对应的分布函数为

$$F_i(x_i) = \frac{\pi}{B(1/2,(s-i)/2)}\int_0^{x_i}(\sin\pi t)^{s-i-1}\mathrm{d}t, 0<i<s.$$

对于一个 $I^{s-1}$ 上有偏差 $D(n)$ 的集合 $\{\boldsymbol{b}_k, k=1,\cdots,n\}$, 我们得到 $I^{s-1}$ 的 $F$- 偏差 $D(n)$ 的一个集合 $\{\boldsymbol{c}_k, k=1,\cdots,n\}$, 此处

$$F_i(c_{ki}) = b_{ki}, i=1,\cdots,s-1, k=1,\cdots,n,$$

最后得到 $S^{s-1}$ 上的一个集合 $P$:

$$\boldsymbol{x}_k = (x_{k1},\cdots,x_{ks})', k=1,\cdots,n \tag{2.14}$$

此处

$$x_{kj} = \prod_{i=1}^{j-1} S_{ki}C_{kj}, j=1,\cdots,s-1, \tag{2.15}$$

$$x_{ks} = \prod_{i=1}^{s-1} S_{ki},$$

其中 $S_{ki} = \sin(\pi c_{ki})$, $C_{ki} = \cos(\pi_k c_i), i=1,\cdots,s-2, S_{k,s-1} = \sin(2\pi c_{k,s-1})$ 及 $C_{k,s-1} = \cos(2\pi c_{k,s-1}), k=1,\cdots,n.$

**例 4** 命 $D$ 为单位单纯形边界的一部分

$$T_{s-1} = \{\boldsymbol{x}: x_1+\cdots+x_s=1, x_i\geqslant 0, i=1,\cdots,s\},$$

它有一个表示

$$x_i = (S_1\cdots S_{i-1}C_i)^2, i=1,\cdots,s-1,$$

$$x_s = (S_1\cdots S_{s-2}S_{s-1})^2,$$

此处 $S_i = \sin(\pi\varphi_i/2), C_i = \cos(\pi\varphi_i/2), i=1,\cdots,s-1$ 及 $\varphi\in I^{s-1}$. 我们有

$$\det(\boldsymbol{TT}') = \left(\pi^{s-1}\prod_{i=1}^{s-1} S_i^{2(s-i)-1}C_i\right)^2 \det(\boldsymbol{SS}'),$$

此处

$$S=\begin{pmatrix} -1 & C_2^2 & S_2^2C_3^2 & \cdots & S_2^2 & \cdots & S_{s-2}^2 & C_{s-1}^2 & S_2^2 & \cdots & S_{s-2}^2 & S_{s-1}^2 \\ 0 & -1 & C_3^2 & \cdots & S_3^2 & \cdots & S_{s-2}^2 & C_{s-1}^2 & S_3^2 & \cdots & S_{s-2}^2 & S_{s-1}^2 \\ \vdots & \vdots & \vdots & \cdots & & & \vdots & & & & \vdots & \\ 0 & 0 & 0 & \cdots & & & -1 & & & & 1 & \end{pmatrix}.$$

注意若将 $\boldsymbol{S}$ 换为 $\boldsymbol{AS}$, 其中 $\boldsymbol{A}$ 为满足 $\det \boldsymbol{A}=\pm 1$ 的 $(s-1)\times(s-1)$ 方阵, 则 $\det(\boldsymbol{SS}')$ 不变. 令证明存在一个 $(s-1)\times(s-1)$ 方阵 $\boldsymbol{A}$ 满足 $\det \boldsymbol{A}=\pm 1$, 使

$$\boldsymbol{AS}=\begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & -1 \end{pmatrix}=\boldsymbol{V}_{s-1},$$

事实上, 若 $s=2$, 则 $S=(-1,1)$, 所以论断成立. 现在假定 $s>2$ 及论断对于 $s-1$ 成立, 则

$$\begin{pmatrix} 1 & -S_2^2 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}\boldsymbol{S}=\begin{pmatrix} -1 & 1 & 0 & \cdots & & 0 & \\ 0 & -1 & C_3^2 & \cdots & S_3^2 & \cdots & S_{s-1}^2 \\ \vdots & \vdots & \vdots & \cdots & & \vdots & \\ 0 & 0 & 0 & \cdots & -1 & 1 & \end{pmatrix}.$$

由归纳法假定可知存在一个 $(s-2)\times(s-2)$ 方阵 $\boldsymbol{A}_1$ 满足 $\det \boldsymbol{A}_1=\pm 1$ 及

$$\boldsymbol{A}_1\begin{pmatrix} -1 & C_3^2 & \cdots & S_3^2 & \cdots & S_{s-1}^2 \\ 0 & -1 & \cdots & & \vdots & \\ \cdots & \vdots & \cdots & & \vdots & \\ 0 & 0 & \cdots & -1 & 1 & \end{pmatrix}=\boldsymbol{V}_{s-2}.$$

所以

$$\begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_1 \end{pmatrix}\begin{pmatrix} 1 & -S_2^2 & \boldsymbol{0} \\ 0 & 1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} \end{pmatrix}\boldsymbol{S}=\boldsymbol{V}_{s-1},$$

从而论断成立. 因此

$$\det(\boldsymbol{SS}')=\det(\boldsymbol{V}_{s-1}\boldsymbol{V}'_{s-1})=\det\begin{pmatrix}2&-1&\cdots&0&0\\-1&2&\cdots&0&0\\\vdots&\vdots&\cdots&\vdots&\vdots\\0&0&\cdots&2&-1\\0&0&\cdots&-1&2\end{pmatrix}=\Delta_{s-1},$$

由于 $\Delta_1=2$ 及 $\Delta_t=2\Delta_{t-1}-\Delta_{t-2}(t>2)$, 所以 $\Delta_{s-1}=s$, 从而

$$J(\boldsymbol{\varphi})=\pi^{s-1}s^{1/2}\prod_{i=1}^{s-1}S_i^{2(s-i)-1}C_i,$$

及

$$v(T_{s-1})=\int_{I^{s-1}}J(\boldsymbol{\varphi})d\boldsymbol{\varphi}=s^{1/2}/(s-1)!.$$

因此

$$f_i(\varphi_i)=(s-i)S_i^{2(s-i)-1}C_i, i=1,\cdots,s-1$$

为 $I$ 上的密度函数, 其对应的分布函数为

$$F_i(x_i)=\int_0^{x_i}f_i(\varphi_i)d\varphi_i=(\sin(\pi x_i/2))^{2(s-i)}, i=1,\cdots,s-1.$$

对于 $I^{s-1}$ 上一个有偏差 $D(n)$ 的集合 $\{\boldsymbol{b}_k,k=1,\cdots,n\}$, 我们得到一个集合 $\{\boldsymbol{c}_k,k=1,\cdots,n\}$, 此处

$$c_{ki}=(2/\pi)\arcsin(b_{ki}^{1/(2s-2i)}), i=1,\cdots,s-1, k=1,\cdots,n.$$

最后, 得 $T_{s-1}$ 上的一个集合 $P$:

$$\boldsymbol{x}_k=(x_{k1},\cdots,x_{ks})', k=1,\cdots,n, \tag{2.16}$$

此处

$$\begin{cases}x_{kj}=\prod_{i=1}^{j-1}b_{ki}^{1/(s-i)}(1-b_{kj}^{1/(s-j)}), j=1,\cdots,s-1,\\ x_{ks}=\prod_{i=1}^{s-1}b_{ki}^{1/(s-i)}, k=1,\cdots,n.\end{cases} \tag{2.17}$$

# §3. 某些应用

本节我们将给出连续多元分布的概率与矩的数值计算的数论方法, 基本求积公式已由 (2.6) 及 (2.7) 给出. 已有一系列方法产生 $\boldsymbol{I}^s$ 上的点集 $\{\boldsymbol{b}_k, k=1,\cdots,n\}$(见华罗庚与王元 [6]). 根据我们的经验, 我们推荐下列程序: 命 $(h_1,\cdots,h_s;n)$ 为一个整矢量, 此处 $h_1=1, 0<h_i<n$ 及 $g$、$c$、$d(h_i,n)=1, i=1,\cdots,s$. 命

$$\boldsymbol{P}_n(k)=(kh_1,\cdots,kh_s)=(q_{k1},\cdots,q_{ks})(\mathrm{mod}\ n), k=1,\cdots,n,$$

此处 $0<q_{ki}\leqslant n$. 置

$$b_{ki}=(2q_{ki}-1)/2n.i=1,\cdots,s,k=1,\cdots,n.$$

则当 $(h_1,\cdots,h_s;n)$ 仔细选取时, $\{\boldsymbol{b}_k\}$ 为 $I_s$ 上的一个有较低偏差的点集, 当 $1<s<19$ 时, 一张 $(h_1,\cdots,h_s;n)$ 的表在 [6] 中作为一个附录.

**例 5** 下面问题来自于合成钢工业 (详细情况见方开泰与吴传义 [3]). 命 $\boldsymbol{x}$ 为一个 $s\times 1$ 矢量, 它表示化学元素在合成钢中的百分比及 $(\boldsymbol{\mu x})=(\boldsymbol{\mu}_1,\cdots,\boldsymbol{\mu}_t)'$ 表示钢的对应质量矢量. $\boldsymbol{\mu x}$ 与 $\boldsymbol{x}$ 间的回归方程为

$$\hat{\boldsymbol{\mu}}(\boldsymbol{x})=\boldsymbol{a}+\boldsymbol{B}\boldsymbol{x},$$

此处 $\boldsymbol{a}$ 与 $\boldsymbol{B}$ 为回归系数的 $t\times 1$ 与 $t\times s$ 矩阵及 $\boldsymbol{x}$ 属于一个立方体 $K=[a_1,b_1]\times\cdots\times[a_s,b_s]$. 假定对于每个 $\boldsymbol{x}\in K$, 我们有 $\boldsymbol{\mu}(\boldsymbol{x})\sim N_t(\boldsymbol{a}+\boldsymbol{B}\boldsymbol{x},\boldsymbol{\Sigma})$, (多元正态分布), 此处 $\boldsymbol{a},\boldsymbol{B}$ 与 $\boldsymbol{\Sigma}$ 能被它们的最小二乘估计所用. 命 $T_i, i=1,\cdots,t$ 为常数使当 $\boldsymbol{\mu}_i>T_i, i=1,\cdots,t$, 就称钢已合规格. 因对应于 $\boldsymbol{x}$, 合金钢合格的概率等于

$$p(\boldsymbol{x})=\int_{T_1}^{\infty}\cdots\int_{T_t}^{\infty}n_t(\boldsymbol{y},\hat{\boldsymbol{\mu}}(\boldsymbol{x}),\boldsymbol{\Sigma})\mathrm{d}\boldsymbol{y}. \tag{3.1}$$

此处 $\boldsymbol{y}=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_t)$ 及 $n_t(\boldsymbol{y},\boldsymbol{\mu}(\boldsymbol{x}),\boldsymbol{\Sigma})$ 为 $N_t(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的密度. 如果我们取适当的 $A_i, i=1,\cdots,t$, 使

$$\begin{aligned}p(\boldsymbol{x})&\cong\int_{T_1}^{A_1}\cdots\int_{T_t}^{A_t}n_t(\boldsymbol{y},(\hat{\boldsymbol{\mu}}\boldsymbol{x}),\boldsymbol{\Sigma})\mathrm{d}\boldsymbol{y}\\&=\prod_{i=1}^{t}(A_i-T_i)\frac{1}{n}\prod_{k=1}^{n}n_t(\boldsymbol{z}_k,\hat{\boldsymbol{\mu}}(\boldsymbol{x}),\boldsymbol{\Sigma}),\end{aligned}\tag{3.2}$$

则积分 (3.1) 可以由 (2.6) 来估计, 此处

$$\boldsymbol{z}_k'=(z_{k1},\cdots,z_{kt})=(T_1+(A_1-T_1)b_{k1},\cdots,T_t+(At-T_t)b_{kt}),$$

$k=1,\cdots,n$ 及 $\{\boldsymbol{b}_k\}$ 为 $I^t$ 上的一个均匀散布点集, 为了说明精密度, 在 (3.2) 中置 $\boldsymbol{\Sigma}=I_5$ 为 $5\times5$ 单位矩阵及 $\mu(\boldsymbol{x})=\mathbf{0}$, $T_i=-\mathbf{1}, A_i=\mathbf{1}, i=\mathbf{1},\cdots,\mathbf{5}$, 则得

$$p=\int_{-1}^{1}\cdots\int_{-1}^{1}n_5(\boldsymbol{y},\mathbf{0},\boldsymbol{I}_5)\mathrm{d}\boldsymbol{y}.$$

由 (3.2) 可得:

表 1 说明一个 5 重积分只用了 1 069 个点即得到 5 位精确度.

**表 1**

| $n$ | $p$ 的近似值 |
|---|---|
| 1069 | 0.148299406 |
| 2129 | 0.148295351 |
| 5003 | 0.148291410 |
| 8191 | 0.148291358 |
| $\infty$ | 0.148291347 |

**例 6** 次序统计量的矩, 命 $X_1,\cdots,X_s$ 为具有分布函数 $F(x)$ 与密度函数 $f(x)$ 的总体的一个样本. 命 $Y_s=X_{(1)}\leqslant Y_{s-1}=X_{(2)}\leqslant\cdots\leqslant Y_1=X_{(s)}$ 为它们的次序统计量. 则 $Y_1,\cdots,Y_s$ 的联合密度为

$$s!\prod_{i=1}^{s}f(y_i), y_s<y_{s-1}<\cdots<y_1.$$

$X_{(i)}, i=1,\cdots,s$ 的阶为 $m_1,\cdots,m_s$ 的混合矩为

$$\mu(m_s,\cdots,m_1)=s!\int_{D^*}\prod_{j=1}^{s}y_j^{m_j}f(y_j)\mathrm{d}v \tag{3.3}$$

此处 $D^*=\{-\infty<y_s<y_{s-1}<\cdots<y_1<\infty\}$. 存在 $a$ 与 $b$ 使

$$P(a<y_s,y_1<b)\cong 1.$$

所以作代换 $z_i=(y_i-a)/(b-a), i=1,\cdots,s$, 则得

$$\mu(m_s,\cdots,m_1)=s!(b-a)^s\int_D\prod_{j=1}^{s}[(a+(b-a)z_j)^{mj}f(a + (b-a)z_j)]\mathrm{d}v.$$

此处 $D$ 如例 1 所定义. 由例 1, (2.6) 与 (2.7), 我们建议下面两个计算 $\mu(m_s,\cdots,n_1)$ 的公式:

$$\mu(m_s,\cdots,m_1)\cong s!(b-a)^s\frac{1}{n}\sum_{k=1}^{n}\prod_{j=1}^{s}\left[t_{kj}^{(m_j+1)-1}f\left(\prod_{i=1}^{j}t_{ki}\right)\right],$$

此处 $t_{kj} = a + (b-a)b_{kj}$ 及 $\{\boldsymbol{b}_k, k=1,\cdots,n\}$ 为 $I^s$ 中的一个均匀散布点列, 及

$$\begin{aligned}&\mu(m_s,\cdots,m_1)\\ &\cong s!(b-a)^s\frac{1}{n}\sum_{k=1}^{n}\prod_{j=1}^{s}[a+(b-a)c_{kj}]^{m_j}f(a+(b-a)c_{kj})\\ &=s!(b-a)^s\frac{1}{n}\sum_{k=1}^{n}\prod_{j=1}^{s}[a+(b-a)b_{kj}^{1/(s-j+1)}]^{m_j}f(a+(b-a)b_{kj}^{1/(s-j+1)})\end{aligned}$$

此处 $\{\boldsymbol{c}_k\}$ 与 $\{\boldsymbol{b}_k\}$ 由例 1 给出.

由于 $I=[0.1]$ 上的均匀分布 $U(0,1)$ 的次序统计量是可以用公式表示的，表 2 中给出的例子将说明精密度.

**表 2　U(0, 1) 的次序统计量的混合矩, s=7**

| $n$ | $E(x_{(1)}X_{(3)}X_{(5)})$ | $E(X_{(1)}X_{(2)}^2X_{(3)}^3)$ |
|---|---|---|
| 418 | 0.03887518 | 0.00326433 |
| 597 | 0.03888518 | 0.00339034 |
| 828 | 0.03888511 | 0.00332116 |
| 1010 | 0.03887286 | 0.00332084 |
| 1220 | 0.03889159 | 0.00325168 |
| $\infty$ | 0.03888889 | 0.00326340 |

**例 7**　在研究混料数据时, Aitchison[1] 于 1986 年引进了所谓加性罗吉斯蒂克正态分布, 命 $T_n$ 表示例 4 中所定义的区域, 其中 $n=N-1$. $T_n$ 中的任何 $\boldsymbol{x}$ 皆称为一个混料. 我们记 $\boldsymbol{x}_{-N}$ 为由 $\boldsymbol{x}$ 的前 $n$ 个支量形成的 $n$ 维矢量. 命

$$\boldsymbol{y} = \log(\boldsymbol{x}_{-N}/X_N) = (\log(X_1/X_N),\cdots,\log(X_n/X_N))'. \tag{3.4}$$

方程 (3.4) 给出由 $T_n$ 至 $R^n$ 的一个一一对应. 若 $\boldsymbol{y}\sim N_n(\boldsymbol{\mu\Sigma})$, 则称随机矢量 $\boldsymbol{x}\in T_n$ 具有一个加性罗吉斯蒂克分布 $AN_n(\boldsymbol{\mu\Sigma})$.

Aitchison 给出 $E(\log(X_i/X_j))$, $E(X_i/X_j)$, $\mathrm{Cov}(\log(X_i/X_j),\ \log(X_k/X_l))$, $\mathrm{Cov}(X_i/X_j, X_k/X_l)$ 的表达式. 但在许多实际问题均常要计算 $E(x_i)$ 与 $\mathrm{Cov}(x_i,x_j)$, 这对他来说似乎有些困难.

$AN_n(\boldsymbol{\mu\Sigma})$ 的密度函数为

$$(2\pi)^{-n/2}(\det\boldsymbol{\Sigma})^{-1/2}\left(\prod_{i=1}^{N}X_i^{-1}\right)\exp\left\{-\frac{1}{2}\left(\log\frac{\boldsymbol{x}_{-N}}{\boldsymbol{x}_N}-\boldsymbol{\mu}\right),\boldsymbol{\Sigma}^{-1}\left(\log\frac{\boldsymbol{x}_{-N}}{\boldsymbol{x}_N}-\boldsymbol{\mu}\right)\right\} \tag{3.5}$$

而 $\boldsymbol{x}$ 的混合矩则为

$$E(X_1^{t_1}\cdots X_N^{t_N}) = (2\pi)^{-n/2}(\det\boldsymbol{\Sigma})^{-1/2}\int_{T_n}\prod_{i=1}^{N}\boldsymbol{x}_t^{t_i-1}$$

$$\exp\left\{-\left(\frac{1}{2}\right)\left[\left(\log\frac{\boldsymbol{x}_{-N}}{X_N}-\boldsymbol{\mu}\right),\boldsymbol{\Sigma}^{-1}\left(\log\frac{\boldsymbol{x}_{-N}}{\boldsymbol{x}_N}-\boldsymbol{\mu}\right)\right]\right\}\mathrm{d}v,$$

此处 $dv$ 为 $T_n$ 的体积元素, 由例 4 可知

$$E(X_1^{t_1}\cdots X_N^{t_N})=C\int_{I^n}\prod_{i=1}^{n}(C_i^{2t_t-1}S_i^{2(t_{i+1}+\cdots+t_N)-3})Q(\varphi)\mathrm{d}\varphi$$

此处 $\varphi\in I^n$, $\mathrm{d}\varphi=\prod_{i=1}^{n}d\varphi_i, C=(2/\pi)^{-n/2}(\det\boldsymbol{\Sigma})^{-1/2}N^{1/2}$, 及

$$Q(\varphi)=\exp\{-(1/2)(\boldsymbol{g}(\varphi)-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{g}(\varphi)-\boldsymbol{\mu})\},$$

其中

$$\begin{aligned}g(\varphi)&=(\log(C_1^2)-\log(S_1^2\cdots S_n^2),\cdots,\log(S_1^2\cdots S_{n-1}^2C_n^2)-\log(S_1^2\cdots S_n^2))',\\&=2\left(\log C_1-\sum_{i=1}^{n}\log S_i,\log C_2-\sum_{i=2}^{n}\log S_i,\cdots,\log C_k-\sum_{i=n}^{n}\log S_i\right)'.\end{aligned}$$

由 (2.6) 或 (2.7), 我们即可得 $AN_n(\boldsymbol{u},\boldsymbol{\Sigma})$ 的任何混合矩的近似值.

**例 8** 我们常碰到方向数据问题, 其中统计量是定义于 $S^{t-1}$ 之上的 (见例 3). 所谓的 Langevin 分布就是 Von Mises 与 Fisher 统计量的推广, 它有密度函数

$$C\exp\{k\mu'\boldsymbol{x}\},\boldsymbol{x}\in S^{s-1},$$

此处 $\mu\in S^{s-1},k>0$ 及 $C$ 是一个正规化后的常数 (见 Mardia[8]). 另一个称之为 Scheidegges-Watson 分布, 其密度函数为

$$C\exp\{k(\mu,\boldsymbol{x})^2\},\boldsymbol{x}\in S^{s-1},$$

此处 $c,k$ 与 $\boldsymbol{\mu}$ 的定义如前所述 (见 Watson[11]).

由例 3、(2.6) 与 (2.7), 我们可以计算这两类分布的概率与矩.

## §4. 最 优 化

本节我们将推广不等式 (1.4) 至前面几节中定义的区域, 然后举例说明 §1 提出的程序是有力的.

我们假定变换 $\boldsymbol{x}=\boldsymbol{x}(\varphi)$ 的奇点集合, 即 $J(\varphi)=0$ 的解的集合, 是 $D$ 的维数 $<t$ 的一个子集, 此处 $x$ 是一个 $s$ 维矢量及 $D$ 的维数为 $t$. 所以对于任何 $\varepsilon>0$, 皆存在仅依赖于 $\varepsilon$ 的一个区域 $\mathfrak{G}$ 及两个常数 $c_1,c_2$ 使 $v(\mathfrak{G})<\varepsilon$ 及

$$c_1<f_i(\varphi_i)<c_2,i=1,\cdots,t,$$

此处 $\varphi \in I^t \setminus \mathfrak{G}$. 则 $dF_i^{-1}(\varphi_i)/d\varphi_i, i=1,\cdots,t$ 在 $I^t \setminus \mathfrak{G}$ 上是正的及有界的.

首先在 $I^t \setminus \mathfrak{G}$ 上取有低偏差 $D(n)$ 的集合 $\boldsymbol{b}_k = (b_{k1},\cdots,b_{kt})', k=1,\cdots,n$. 则我们已经证明

$$\boldsymbol{c}_k = \boldsymbol{F}^{-1}(\boldsymbol{b}_k) = (F_1^{-1}(b_{k1}),\cdots,F_t^{-1}(b_{kt}))', k=1,\cdots,n$$

为 $I^t$ 的具有 $F_-$ 偏差 $D(n)$ 的集合. 最后得到 $D$ 的一个集合:

$$\boldsymbol{x}_k = \boldsymbol{x}(\boldsymbol{c}_k), k=1,\cdots,n.$$

命 $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{\varphi}), \boldsymbol{x}^* = \boldsymbol{x}(\boldsymbol{\varphi}^*)$, $\boldsymbol{\varphi} = \boldsymbol{F}^{-1}(\boldsymbol{\psi}), \boldsymbol{\varphi}^* = \boldsymbol{F}^{-1}(\boldsymbol{\psi}^*)$. $\mathrm{d}\boldsymbol{x} = (\mathrm{d}x_1,\cdots,\mathrm{d}x_s)'$, $\mathrm{d}\boldsymbol{\varphi} = (\mathrm{d}\varphi_1,\cdots,\mathrm{d}\varphi_t)'$, $\mathrm{d}\psi = (\mathrm{d}\psi_1,\cdots,\mathrm{d}\psi_t)'$, 及 $S$ 表示对角方阵

$$\boldsymbol{S} = \mathrm{diag}(\mathrm{d}\varphi_1/\mathrm{d}\psi_1,\cdots,\mathrm{d}\varphi_t/\mathrm{d}\psi_t).$$

则

$$\mathrm{d}\boldsymbol{x}'d\boldsymbol{x} = \mathrm{d}\boldsymbol{\varphi}'\boldsymbol{T}\boldsymbol{T}\mathrm{d}\boldsymbol{\varphi} = \mathrm{d}\boldsymbol{\psi}'\boldsymbol{S}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{S}\mathrm{d}\boldsymbol{\psi}.$$

由于 $S$ 与 $T$ 在 $I^t \setminus \mathfrak{G}$ 上有界, 所以

$$\begin{aligned}\mathrm{d}(\boldsymbol{x},\boldsymbol{x}^*) &= \int_{x(\varphi)}^{x(\varphi^*)} (\mathrm{d}\boldsymbol{x}'\mathrm{d}\boldsymbol{x})^{1/2} \\ &= \int_{\psi}^{\psi^*} (\mathrm{d}\psi'\boldsymbol{S}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{S}\mathrm{d}\psi)^{1/2} < c(\varepsilon)\int_{\psi}^{\psi^*} (\mathrm{d}\psi'\mathrm{d}\psi)^{1/2} \\ &= c(\boldsymbol{\varepsilon})\mathrm{d}_t(\boldsymbol{\psi},\boldsymbol{\psi}^*),\end{aligned}$$

此处 $d_t(\boldsymbol{y},\boldsymbol{z})$ 表示 $R^t$ 中的欧氏距离. 因此由 (1.4) 可得

$$d(n,D) = \max_{x\in D}\min_{1\leqslant k\leqslant n} d(\boldsymbol{x},\boldsymbol{x}_k) \leqslant 2t^{1/2}c(\boldsymbol{\varepsilon})D(n)^{1/t}, \tag{4.1}$$

所以若 $\boldsymbol{x}_k, k=1,\cdots,n$ 在 $D$ 上均匀散布, 则函数在这些点上的最大值可以作为函数在 $D$ 上的全局极大的近似值.

**例 9** 加性罗吉斯蒂克椭球分布. 定义于 $T_{s-1}$ 上的所谓加性罗吉斯蒂克椭球分布是例 7 中定义的加性罗吉斯蒂克正态分布的推广, 它有密度函数

$$f(\boldsymbol{x}) = (\det\boldsymbol{\Sigma})^{-1/2}\prod_{i=1}^{\infty} x_i^{-1}g(\boldsymbol{x}) \tag{4.2}$$

此处

$$g(\boldsymbol{x}) = g\left(\left(\log\frac{\boldsymbol{x}_{-s}}{\boldsymbol{x}_{\ s}} - \boldsymbol{\mu}\right)'\boldsymbol{\Sigma}^{-1}\left(\log\frac{\boldsymbol{x}_{-s}}{\boldsymbol{x}_s} - \boldsymbol{\mu}\right)\right) \tag{4.3}$$

及 $\boldsymbol{x}_{-s}$ 为 $\boldsymbol{x}$ 的前 $s-1$ 个分量构成的 $(s-1)$- 维矢量. $f(\boldsymbol{x})$ 的重数尚无解析表示. 但我们用 $T_{s-1}$ 上的均匀散布点集来求出重数的近似值.

当 (4.3) 中的函数 $g$ 取形式

$$g(u)=C(1+u/m)^{-p}, p>s/2, m>0, \tag{4.4}$$

此处

$$C=(\pi m)^{-s/2}\Gamma(p)/\Gamma(p-s/2),$$

则对应的分布就称为加性罗吉斯蒂克椭球 Pearson 型 VII 分布. 在这情形下, 寻求 $f(\boldsymbol{x})$ 的重数等价于求得函数

$$h(\boldsymbol{x})=\prod_{i=1}^{s}\boldsymbol{x}_i^{-1}\left[1+\left(\log\frac{\boldsymbol{x}_{-s}}{\boldsymbol{x}_s}-\mu\right)\boldsymbol{\Sigma}^{-1}\left(\log\frac{\boldsymbol{x}_{-s}}{\boldsymbol{x}_s}-\mu\right)/m\right]^{-p}$$

在 $T_{s-1}$ 上的最大值. 表 3 中的结果表示在情况 $s=3, p=9, m=5.5$ 及

$$\mu=\begin{pmatrix}0\\0\end{pmatrix},\quad \boldsymbol{\Sigma}=\begin{pmatrix}1 & -0.7\\ -0.7 & 1\end{pmatrix}$$

时, 重数的近似值很接近于重数 (1/2, 1/3, 1/3)′.

**表 3　重数的近似值**

| $n$ | $M_n$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|
| 233 | 26.55203 | 0.3271035 | 0.3306723 | 0.3422242 |
| 377 | 26.82553 | 0.3061467 | 0.3598099 | 0.3340434 |
| 610 | 26.48600 | 0.3604561 | 0.3150540 | 0.3244899 |
| 4181 | 26.89095 | 0.3256147 | 0.3368701 | 0.3375152 |
| 10946 | 26.99783 | 0.3334971 | 0.3337690 | 0.3827339 |
| 17711 | 26.98296 | 0.3292827 | 0.3388427 | 0.3318746 |

注意: 在表 3 中, 一般说来当 $n$ 增加时, $M_n$ 亦增加. 但有时当 $n>n'$ 时反而有 $M_n<M_n'$. 这是由于对于不同的 $n$, $\{c_k\}$ 可以是完全不同的集合. 因此我们建议用序贯的方法来改进上述结果.

下面的程序是为这一问题而设计的.

第 1 步: 在 $D_0=T_{s-1}$ 上选取一个适当的均匀散布点集 $\{\boldsymbol{x}_k, k=1,\cdots,n_0\}$. 找出函数在这些点的最大值 $M_0$, 并假定函数在 $\boldsymbol{x}_0^*=(x_{01},\cdots,x_{0s})'$ 处达到最大值.

第 2 步: 找出 $D_0$ 的子区域 $D_1\subset D_0$ 并使 $\boldsymbol{x}_0^*\subset D_1$. 例如 $D_1$ 为一个区域, 其中 $\boldsymbol{x}_0^*$ 位于 $D_1$ 的重心附近. 我们可以取 $a_i, i=1,\cdots,s$ 使

$$0\leqslant a_i<x_{0i}, i=1,\cdots,s.$$

置 $a = a_1 + \cdots + a_s$ 及 $b_i = a_i + 1 - a, i = 1, \cdots, s$ 则

$$1 \geqslant b_i \geqslant a_i + \sum_{j=1}^{s} x_{0j} - \sum_{k=1}^{s} a_k = x_{0i} + \sum_{\substack{j=1 \\ i \neq j}}^{s} (x_{0j} - a_j) \geqslant x_{oi},$$

$i = 1, \cdots, s$. 记

$$D_1 = \{x = (x_1, \cdots, x_s)' : a_i \leqslant x_i \leqslant b_i, i = 1, \cdots, s, x \in D_0\}.$$

命 $z_k, k = 1, \cdots, n_1$, 为 $T_{s-1}$ 上的一个均匀散布点集. 则得到一个集合 $\{x_k, k = 1, \cdots, n_1\}$, 此处

$$x_{ki} = a_i + (1-a) z_{ki}, i = 1, \cdots, s, k = 1, \cdots, n_1,$$

它在 $D_1$ 上均匀散布. 命 $M_1$ 为函数在这些点的最大值. 它由 $x_1^*$ 达到.

第 3 步: 假定在第 $j$ 步, 我们找到了函数的最大值 $M_j$ 及对应的点 $x_j^*$, 用类似的方法我们可以将 $D_j$ 缩小为 $D_{j+1}$. 在 $D_{j+1}$ 上找一个点集并由此得到函数在这个集合上的最大值 $M_{j+1}$ 及对应的点 $x_{j+1}^*$.

重复步骤 3 直至搜索区域较小为止. 最后一步得到的极大值 $M_{j+1}$ 可望很接近于函数的全局极大 $M$.

将上述程序用于我们的问题: 每一步均取 $n_0 = n_1 = \cdots = 233$, 结果列于表 4 之中, 它改进了表 3 的结果.

表 4 最优化的序贯方法

| 序号 | $a_i$ | $b_i$ | $M_i$ | $x_1^*$ | $x_2^*$ | $x_3^*$ |
|---|---|---|---|---|---|---|
| 1 | 0.0000 | 1.0000 | 26.55203 | 0.3271035 | 0.3306723 | 0.3422242 |
| 2 | 0.3000 | 0.4000 | 26.97836 | 0.3304331 | 0.3343395 | 0.3352274 |
| 3 | 0.3300 | 0.3400 | 26.99543 | 0.3327104 | 0.3330673 | 0.3342224 |
| 4 | 0.3330 | 0.3340 | 26.99994 | 0.3332711 | 0.3333067 | 0.3334223 |
| | 全局极大 | | 27.00000 | 0.3333333 | 0.3333333 | 0.3333333 |

我们做了许多例子都表明上述序贯方法是有利的.

## 参 考 文 献


[1] Aitchison, J., The statistical Analysis of Compositional Data, Chapman and Hall, Loodon/New York, (1986).

[2] Avriel, M., Nonlinear Programming, Analysis and Methods, Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1976.

[3] 方开泰与吴传义 (Fnag, K. T. & Wu. C. Y)., The extrime value problem of some probability function, *Acta Math. Appl. Sinica*, **2** (1979), 132~148.

[4] Hlawka, E., Funktionen von beschrankter Variation in der Theorie Gleichverteilnug, *Ann. Mat. pure Appl.,* **54**, (1961), 325~ 333.

[5] Hlawka, E. & Much, R., A transformations of equidistributed sequences, in "Applications of Number Theory to Numerical Analysis" (Zaremba. S. K. ed). Acad. Press. New York, 1972, 371~388.

[6] 华罗庚与王元 (Hua, L. K. & Wang. Y.)., Applications of Number Theory to Numerical Analysis, Springer-Verlag (Heidelberg) and Science Press (Beijing), 1981.

[7] Koksma, I, F., Een algemeene stelling uit de theorie der gelijkmatige Verdeeling modulo 1, *Math. B (Zutphen)*, **11**, (1942~1943), 7~11.

[8] Madia, K. V., Statistics of Directional Data, Acad. Press, New York, (1972).

[9] Niederreiter, H., Metric theorems on the distribution of sequenes, *Proc. Symp. Pure Math.,* **24**, *AMS Pro. R. I.,* (1973), 195~212.

[10] Niederreiter, H., A quasi-Monte Carlo method for the approximate computation of the extreme values of a function, *Studies in Pure Math, Birkhauser, Basel.* (1983), 523~529.

[11] Watson, G. S., Statistics on Sphere, Wiley, New York, (1983).

[12] Weyl, H., Uber die Gleichverteilung der Zahlen mod Eins, *Math. Ann.,* **77**. (1916), 313~ 352.

[13] Zielinski, R., On the Monte Carlo evaluaton of the extreme values of a function, *Algorytncy*, **2** (1965), 7~13.

# 应用统计中的数论方法 (II)*

王元 方开泰
(中国科学院数学研究所) (中国科学院应用数学研究所)

## 提 要

在本文中, 作者将给出上文定义的 $F$- 均匀散布贯在试验设计、混料试验、几何概率与模拟中的应用.

## §1. 导 言

我们在前文里建议了一个产生 $R^s$ 中一个区域 $D$ 上均匀散布点集的方法, 并给出了它们对连续多元分布的概率与矩的近似计算及最优化问题的应用. 本文我们将这类均匀散布点列用于独立因素及混料试验设计 (见 §2 与 §3), 和几何概率问题 (见 §4). 我们还将举例对数论方法与其他方法加以比较.

前文 [13] 中引用的定义与记号, 在此均保留.

## §2. 均 匀 设 计

如果有 $s$ 个因素及每个因素有 $n$ 个水平, 则所有可能的试验总和为 $n^s$, 基于正交拉丁方理论与群论的正交设计是从这些试验中选出 $O(n^2)$ 个试验. 若 $n$ 比较大, 则正交设计的试验次数仍嫌太多. BIB(平衡不完全区组设计) 法可以降低试验次数, 但仅限于 $s=2$. 因此仍需寻找一个降低试验次数的方法.

基于数论方法, 王元与方开泰于 1981 年建立了一类试验设计方法, 即均匀设计, 它已在中国满意地用于纺织工业、冶金工业、机械工业与农业中一些新产品的试验设计.

我们提供一系列均匀设计表 $U_n(b^c)$, 此处 $n$ 表示试验次数, $b$ 表示水平数及 $c$ 表示因素的最大个数. 例如在一项试验设计中共有三个因素 $A, B, C$, 而每个因素有 11 个水平 $A_i, B_j, C_k$, 则可以利用表 1 的 $U_{11}(11^6)$.

* 原载 "数学年刊", 11B, 3, 1990, 384~394.
这一工作受到中国国家自然科学基金及中国科学院的基金支持.

**表 1　$U_{11}(11^6)$**

| 序号 \ 列 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 5 | 7 | 10 |
| 2 | 2 | 4 | 6 | 10 | 3 | 9 |
| 3 | 3 | 6 | 9 | 4 | 10 | 8 |
| 4 | 4 | 8 | 1 | 9 | 6 | 7 |
| 5 | 5 | 10 | 4 | 3 | 2 | 6 |
| 6 | 6 | 1 | 7 | 8 | 9 | 5 |
| 7 | 7 | 3 | 10 | 2 | 5 | 4 |
| 8 | 8 | 5 | 2 | 7 | 1 | 3 |
| 9 | 9 | 7 | 5 | 1 | 8 | 2 |
| 10 | 10 | 9 | 8 | 6 | 4 | 1 |
| 11 | 11 | 11 | 11 | 11 | 11 | 11 |

附着于每一张表 $U_n(b^c)$, 我们还有另一张表, 它指出对于 $s$ 个因素的试验应选取哪些列. 附着于 $U_{11}(11^6)$ 的表见表 2.

**表 2　附着于 $U_{11}(11^6)$ 的表**

| 因素个数 | 推荐列数 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | | | | | 1 | | 5 | | | | |
| 3 | | | | 1 | | 4 | | 5 | | | |
| 4 | | | 1 | | 2 | | 4 | | 5 | | |
| 5 | | 1 | | 2 | | 3 | | 4 | | 5 | |
| 6 | 1 | | 2 | | 3 | | 4 | | 5 | | 6 |

对于我们的问题, 应该推荐使用表 1 中的 1, 4, 5, 列. 最后我们将试验设计列于表 3 之中. 所以对于 3 个因素及每个因素皆有 11 个水平的试验, 我们仅安排了 11 次试验.

均匀设计表是由一个整矢量 $(h_1, \cdots, h_s, n)$ 生成的, 此处 $h_1 = 1, h_1 < h_2 < \cdots < h_s$ 及 $g, c, d(h_i, n) = 1, i = 1, \cdots, s$. 命

$$P_n(k) = (kh_1, \cdots, kh_s) \equiv (q_{k1}, \cdots, q_{ks}) \pmod{n}, \tag{2.1}$$

此处 $0 < q_{kj} \leqslant n, k = 1, \cdots, n, j = 1, \cdots, s$. 表 $U_n(n^s)$ 就是 $(q_{kj})$ 生成的. 当 $n = 11$, $s = 6, h_1 = 1, h_2 = 2, h_3 = 3, h_4 = 5, h_5 = 7$ 及 $h_6 = 10$, 均匀设计表就是 $U_{11}(11^6)$(见表 1).

因 $1 \leqslant h_i < n$ 及 $g$、$c$、$d$. $(h_i, n) = 1, i = 1, \cdots, s$, 所以这种 $h_i$ 的个数等于 Euler 函数 $\varphi(n)$,

$$\varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right), \tag{2.2}$$

此处 $p$ 过 $n$ 的所有素因子. 由于 $-h \equiv n-h(\bmod n)$, 所以矩阵 $(q_{kj})$ 的秩, 当 $n>2$ 时, 不超过 $1+\varphi(n)/2$; 即因素个数必须 $\leqslant \varphi(n)/2+1$. 又由于 $h_1=1$, 所以 $h=(h_1,\cdots,h_s)'$ 最多只有 $\binom{\varphi(n)/2}{s-1}$ 个选择. 我们希望从这些 $h$ 中找出 "最好" 的 $h$ 来. 王元与方开泰 (1981) 指出 "最好" 的 $h$ 是使

$$D(\boldsymbol{h})=\frac{1}{n}\sum_{k=1}^{n}\sum_{v=1}^{s}\left(1-\frac{2}{\pi}\log\left(2\sin\left\{\frac{\pi k h_v}{n+1}\right\}\right)\right) \tag{2.3}$$

达到最小的 $\boldsymbol{h}$, 此处 $\{x\}$ 表示 $x$ 的分数部分, 这个 $h$ 所对应的点集 (2.1) 就是均匀设计.

**表 3 试验设计**

| 试验序号 \ 因素 | A | B | C |
|---|---|---|---|
| 1 | $A_1$ | $B_5$ | $C_7$ |
| 2 | $A_2$ | $B_{10}$ | $C_3$ |
| 3 | $A_3$ | $B_4$ | $C_{10}$ |
| 4 | $A_4$ | $B_9$ | $C_6$ |
| 5 | $A_5$ | $B_3$ | $C_2$ |
| 6 | $A_6$ | $B_8$ | $C_9$ |
| 7 | $A_7$ | $B_2$ | $C_5$ |
| 8 | $A_8$ | $B_7$ | $C_1$ |
| 9 | $A_9$ | $B_1$ | $C_8$ |
| 10 | $A_{10}$ | $B_6$ | $C_4$ |
| 11 | $A_{11}$ | $B_{11}$ | $C_{11}$ |

当 $n$ 较大时, 需很大的计算量才能得到这样的 $h$. 为此我们建议用下面形状的点集

$$Q_n(k)=(k,kb,\cdots,kb^{s-1}) \pmod n, 1\leqslant k\leqslant n \tag{2.4}$$

来代替 $P_n(k)$, 此处 $b$ 为满足 $1<b\leqslant n/2$ 及 $b^i\not\equiv b^j \pmod n$, $1\leqslant i<j\leqslant s-1$ 的整数. 当 $n$ 为素数时, 则通常取 $b$ 为 mod $n$ 的一个原根. 我们亦称使 $D(b)$ 取极小的 $\boldsymbol{b}=(1,b,\cdots,b^{s-1})'$ 构成的点集 (2.4) 为一个均匀设计. 绝大多数均匀设计表都是由形如 (2.4) 的点集产生的.

偶数 $n$ 所对应的均匀设计表可以由 $n+1$ 对应的表中去掉最后一行来求得. 例如表 $U_{10}(10^6)$ 即可以由去掉 $U_{11}(11^6)$ 的最后一行以得到 (见附 1).

由于试验次数相比于因素与水平个数显得太小, 所以由均匀设计得到的数据不能用通常的方差分析方法来进行分析, 但我们可以用回归分析或逐步回归分析方法来处理数据.

**例 1** 一种维尼纶设计中需考虑以下因素：

A：温度 ($C$);

B：时间 ($m$);

C：木醇浓度 ($g/l$);

D：硫酸浓度 ($g/l$);

E：芒硝浓度 ($g/l$).

每个因素都有 7 个水平, 如表 4 所示：

**表 4 因素与水平**

| 因素 \ 水平 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| A | 64 | 66 | 68 | 70 | 72 | 74 | 76 |
| B | 14 | 16 | 18 | 20 | 22 | 24 | 26 |
| C | 18 | 20 | 22 | 24 | 26 | 28 | 30 |
| D | 206 | 212 | 218 | 224 | 230 | 236 | 242 |
| E | 70 | 70 | 85 | 85 | 85 | 100 | 100 |

如果用正交设计, 则需安排 49 次实验, 从而得到下面的线性回归方程

$$\hat{y} = -42.37 + 0.55x_1 + 0.38x_2 + 0.26x_3 + 0.10x_4 - 0.04x_5. \qquad (2.5)$$

其多重相关系数为 $R = 0.97$, 而标准离差为 $\hat{\sigma} = 0.83$, 此处 $\hat{y}$ 表示维尼纶的质量. 现在我们来用均匀设计并选择表 $U_{14}(14^5)$. 这是 14 个水平的用表. 我们用拟水平方法, 即将原来的水平重复一次, 于是得回归方程如下：

$$\hat{y} = -57.97 + 0.37x_1 + 0.46x_2 + 0.38x_3 + 0.17x_4 + 0.04x_5,$$

其中 $R = 0.96$ 及 $\sigma = 1.13$. 方程 (2.6) 与 (2.5) 很接近, 所以虽然只做了 14 次试验, 但结果并不错.

## §3. 混 料 试 验

若 $s$ 个因素 $X_1, \cdots, X_s$ 均取非负值且适合 $X_1 + \cdots + X_s = 1$, 这种实验就称为混料试验. 在化工与冶金产品的设计中, 常常需做这样的试验. 近 20 年来, 有许多统计文献是讨论这个问题的：Scheffe (1958) 引入了单纯形一格点设计及多项式模型. 1963 年, 他又建立了单纯形一重心设计. Draper 与 Lawrence (1965) 建立了一个设计, 它通过试验区域的回报的最小二乘估计来达到预定模型的参偏相合与离差极小. Thompson 与 Myers (1968) 建立的设计是基于一个椭球区域置于单纯形因素空间中、Cornell (1975) 建立了轴设计并给予这个分支以全面总结 (1973, 1981).

本节我们将提出一个使用单纯形上的均匀散布点集来处理混料试验的方法. 它与 §2 提出的均匀设计有同样的优越性. 注意到混料的陈述, 所以称这种设计为混料均匀设计 (UDEM). 命

$$T_{s-1} = \{(x_1, \cdots, x_s) : x_i \geqslant 0, i = 1, \cdots, s, x_1 + \cdots + x_s = 1\}$$

为 $R^s$ 中单位单纯形的边界的一部分. UDEM 的想法为在 $T_{s-1}$ 上一个均匀散布的 $n$ 个点上来安排试验.

命 $\{q_{ki}, i = 1, \cdots, s-1, k = 1, \cdots, n\}$ 为由 (2.1) 定义的一个均匀设计. 则 $\{\boldsymbol{b}_{ki}\}$ 为 $I^{s-1}$ 上的一个均匀散布点集, 其中

$$b_{ki} = \frac{2q_{ki} - 1}{2n}, k = 1, \cdots, n, i = 1, \cdots, s-1 \tag{3.1}$$

在 [13]、§2 中, 我们建立了一个产生 $T_{s-1}$ 上均匀散布的点集 $\{P_k, k = 1, \cdots, n\}$ 的方法, 其中 $P_k = \{x_{k1}, \cdots, x_{ks}\}'$ 及

$$\begin{cases} x_{kj} = \prod_{i=1}^{j-1} b_{ki}^{1/(s-i)} (1 - b_{kj}^{1/(s-j)}), & j = 1, \cdots, s-1 \\ x_{ks} = \prod_{i=1}^{s-1} b_{ki}^{1/(s-i)}, & k = 1, \cdots n. \end{cases} \tag{3.2}$$

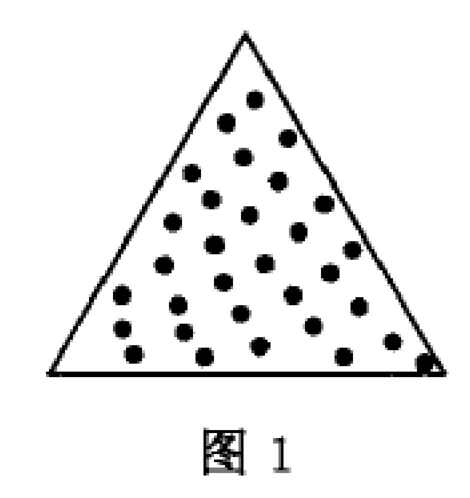


图 1

当 $s = 3$ 及 $n = 31$ 时, 集合 (3.2) 的均匀性由图 1 所示.

**例 2**　考虑回归模型

$$Y = e + \sum_{i=1}^{s} e_i X_i + \sum_{i,j=1}^{s} e_{ij} X_i X_j + \varepsilon,$$

此处 $\varepsilon$ 表示随机误差. 由于 $X_1 + \cdots + X_s = 1$, 所以模型可以化为

$$Y = e + \sum_{i=1}^{s-1} e_i X_i + \sum_{i,j=1}^{s-1} e_{ij} X_i X_j + \varepsilon. \tag{3.3}$$

研究特殊模型

$$Y = X_1 + X_2 - 3X_1^2 - 3X_2^2 + X_1 X_2 + \varepsilon, \tag{3.4}$$

此处 $\varepsilon \sim N(o, \sigma^2)$. 当 $\sigma$ 较小时 (例如 $\sigma = 0.005$), 我们用 UDEM(3.2) 得到 $n = 17$ 及 $s = 3$ 时的数据 (见表 5).

对应的回归模型为

$$\hat{Y} = -0.0376 + 1.1162X_1 + 1.1197X_2 - 3.0842X_1^2$$

$$- 3.0880X_2^2 + 0.8336X_1X_2, \tag{3.5}$$

它接近于模型 (3.4). 方程 (3.5) 的多重相关系数为 $R = 0.9999$ 及标准离差估计为 $\sigma = 0.0054$, 它亦接近于 $\sigma = 0.005$.

当 $\sigma$ 变大时, 我们得不到这样好的结果. 例如, 考虑模型

$$Y = 10 + X_1 - 3X_1^2 - 3X_2^2 + X_1X_2 + \varepsilon, \tag{3.6}$$

此处 $\varepsilon \sim N(0, \sigma^2)$, 其中 $\sigma = 0.3$. 由 (3.2) 安排 15 个试验, 数据列于表 6.

对应的回归方程为

$$\hat{Y} = 10.0908 + 0.7972X_1 - 3.4542X_1^2 - 2.6733X_2^2 + 0.8884X_1X_2, \tag{3.7}$$

其中 $R = 0.9003$ 及 $\hat{\sigma} = 0.2891$. 注意由于 $X_1$ 与 $X_1^2$ 之间有高相关性, 所以 (3.7) 偏离了模型 (3.6), 原来的模型在 $X_1 = 0.171$ 及 $X_2 = 0.0286$ 时, 回报 $Y$ 达到极大值 10.0857. 由 (3.7) 易知当 $X_1 = 0.105$, $X_2 = 0.0196$ 时, $\hat{Y}$ 达极大值 10.0728, 接近于 10.0857. 因此由混料的最佳表达的观点看, 这一结果似乎是不错的.

**表 5 数据**

| 序号 | $X_1$ | $X_2$ | $Y$ |
|---|---|---|---|
| 1 | .829 | .076 | −1.100 |
| 2 | .703 | .253 | −.541 |
| 3 | .617 | .102 | −.391 |
| 4 | .546 | .307 | −.157 |
| 5 | .486 | .045 | −.160 |
| 6 | .431 | .284 | .038 |
| 7 | .382 | .546 | −.230 |
| 8 | .336 | .215 | .146 |
| 9 | .293 | .520 | −.103 |
| 10 | .252 | .110 | .163 |
| 11 | .214 | .439 | .031 |
| 12 | .178 | .798 | −.889 |
| 13 | .143 | .328 | .134 |
| 14 | .109 | .708 | −.644 |
| 15 | .076 | .190 | .155 |
| 16 | .045 | 590 | −.388 |
| 17 | .015 | .029 | .000 |

表 6　数据

| 序号 | $X_1$ | $X_2$ | $Y$ |
|---|---|---|---|
| 1 | 0.817 | 0.055 | 8.508 |
| 2 | 0.684 | 0.179 | 9.464 |
| 3 | 0.592 | 0.340 | 9.935 |
| 4 | 0.517 | 0.048 | 9.400 |
| 5 | 0.452 | 0.201 | 10.680 |
| 6 | 0.394 | 0.384 | 9.748 |
| 7 | 0.342 | 0.592 | 9.698 |
| 8 | 0.293 | 0.118 | 10.238 |
| 9 | 0.247 | 0.326 | 9.809 |
| 10 | 0.204 | 0.557 | 9.732 |
| 11 | 0.163 | 0.809 | 8.933 |
| 12 | 0.124 | 0.204 | 9.971 |
| 13 | 0.087 | 0.456 | 9.881 |
| 14 | 0.051 | 0.727 | 8.892 |
| 15 | 0.017 | 0.033 | 10.139 |

## §4.　几何概率与模拟

本节. 我们用两个“情况研究”方法来说明 $D$ 上均匀散布点集在几何概率问题与模拟上的应用. 读者不难从这两个例子中去了解一般的原则. 这里的问题是在实际中较长时间得不到满意解答者. 现在我们用数论方法建立一个程序来发现它们的可行解.

A. 一个固定圆与一系列随机圆的公共部分的面积, 给予一个以原点为中心的单位圆 $K$. 假定有 $m$ 个随机圆 $O_1,\cdots,O_m$, 其中心与半径分别为 $P_1,\cdots,P_m$ 与 $R_1,\cdots,R_m$. 又假定

$$P_i\sim N_2(\mathbf{0},\sigma_i^2\boldsymbol{I}_2),$$

此处 $\sigma_i>0$ 及 $I_2$ 为 $2\times 2$ 单位方阵. 命 $S$ 为 $K$ 与这些圆的并的公共部分, 即

$$S=K\cap(O_1\cup\cdots\cup O_m).$$

(见图 2)$S$ 的面积亦记为 $S$. 需寻求 $S$ 的分布. 因两个圆的公共部分的面积易于用这两个圆的联心线的长度来表示, 所以当 $m=1$ 时, 易找出 $S$ 的分布. 当 $m>1$, 则很难找出一个可行的方法来求得 $S$ 的分布. 一个自然的方法为使用模拟. 经典的方法就是所谓格子点方法. 命 $ABCD$ 为单位圆 $K$ 的外接正方形, 如图 3 所示. 将 $ABCD$ 分割成边长为 $2/(n-1)$ 的 $n^2$ 个相等的小正方形, 则得 $n^2$ 个格子点

$$\left(-1+\frac{2i}{n-1},-1+\frac{2j}{n-1}\right),\quad 0\leqslant i,\quad j\leqslant n-1.$$

假定其中有 $N$ 个点落在 $K$ 之中. 我们现在用计算机产生 $m$ 个随机圆, 其中心为 $P_i \sim N_2(0, \sigma_i^2 I_2)$ 及半径为 $R_i(i = 1, \cdots, m)$. 假定这 $N$ 个格点中有 $M$ 点被这 $m$ 个随机圆所覆盖. 则得到 $S$ 的一个观测值 $\pi M/N$. 然后再生成 $m$ 个随机圆并得到另一个观测值. 继续这一步骤, 我们得到 $S$ 的一个经验分布. 我们称这个方法为方法 I. 它的收敛速是很慢的. 因为有 $(O\sqrt{N})$ 个点分布 $K$ 的边界附近, 所以即使取 $N$ 很大, 收敛速度仍很慢, 这是颇严重的事.

但我们可以用 $s = 2$ 时的点集 (3.1)(经线性变换) 来代替上述 $n^2$ 个格点来进行模拟 (见图 1). 我们称这一方法为方法 II. 它比方法 I 有较快的收敛速度及较高的精密度. 例如取 $m = 1$, 我们可以将这两个方法算出的 $S$ 的值与 $S$ 的真值相比较. 用计算机 IBMPC/XT, 第一个方法用大小为 1000 的样本, 耗时 180 分钟, 所得误差为 0.15, 第二个方法用大小为 1500 的样本, 耗时 4 分钟, 所得误差为 0.02. 一般来说, 欲得同样精密的结果, 方法 II 比方法 I 约快 100~1000 倍.

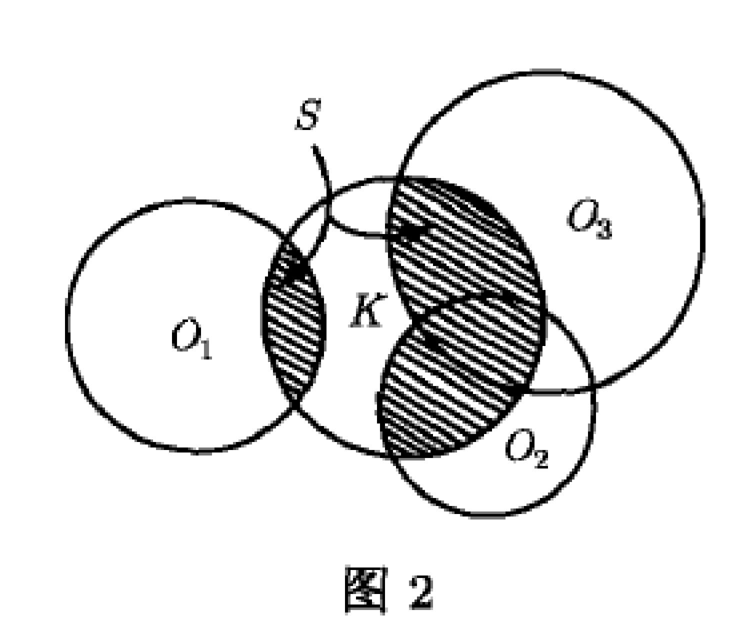



图 2

注意 $s = 2$ 时的点集 (3.1) 是定义于 $ABCD$ 之中, 而不是定义于 $K$ 之中. 受到 $K$ 上数值积分法的启发 (见 [13]), 我们可以用 (3.1) 得到 $K$ 上的一个均匀散布点集. 命

$$\begin{cases} x = r\cos(2\pi\theta), \\ y = r\sin(2\pi\theta), \quad 0 \leqslant \theta < 2\pi, \quad 0 \leqslant r \leqslant 1, \end{cases} \tag{4.1}$$

及 $(\theta_i, r_i)(1 \leqslant i \leqslant N)$ 为一个点集 (4.1). 若集合的每一点

$$Q_n(i) = (r_i \cos 2\pi\theta_i, r_i \sin 2\pi\theta_i), 1 \leqslant i \leqslant N \tag{4.2}$$

均有一个适当的 "权"$w_i$, 则 (4.2) 可以在 $K$ 中均匀散布 (见图 4). 由于变换 (4.1) 有 Jacobian $2\pi r$, 所以我们定义 $Q_n(i)$ 的权为 $2\pi r_i$. 因此若有 $M$ 个点 $Q_n(i_j)(j = 1, \cdots, M)$ 被这 $m$ 个随机圆所覆盖, 则

$$2\pi \sum_{j=1}^{M} r_{i_j}$$

可以作为 $K$ 被 $m$ 个随机圆覆盖部分的面积的近似值. 我们称这个方法为方法 III. 为了比较方法 II 与 III, 我们每取 $m = 1$ 并考虑两个单位圆 $K$ 与 $O$, 它们的中心距离为 $d$(见图 5). 对于方法 II, 我们用形如 (3.1) 的 1069 个点的点集, 其中有 844 个点落于 $K$ 中, 对于方法 III, 我们一个有 823 个点的形如 (3.1) 的点集, 结果列于表 7 之中

表 7

| 误差 \ d | 0.1 | 0.75 | 0.8 | 1.3 |
|---|---|---|---|---|
| 方法 II | −0.55% | 0.07% | 0.19% | 0.013% |
| 方法 III | 0.00% | 0.04% | 0.12% | 0.08% |

由 [13] 之例 2, 我们可以有一个 $K$ 上均匀散布的点集 $\{\boldsymbol{x}_k, k=1,\cdots,n\}$, 由此可得另一个方法 —— 方法 IV, 现在可以仿照方法 II, 用点集 $\{\boldsymbol{x}_k, k=1,\cdots,n\}$ 来进行模拟. 我们的模拟结果表明方法 III、IV 有同等精密度.

B. 用固定宽度的随机带来覆盖球面.

这个问题来自滚动扎钢 [1], 人们希望用随机球形扎滚来代替固定轧滚, 以延长轧滚寿命. 这一问题的数学模型如下: 命 $S$ 表示单位球 $X_1^2+X_2^2+X_3^2=1$ 及 $\delta$ 表示一个适合 $0<\delta<0.3$ 的常数. 命 $R$ 表示一个大圆, 它在 $S$ 上均匀散布及 $G_\delta(R)$ 为 $S$ 上一个宽度为 $\delta$ 并以 $R$ 为等分线的带子, 命 $G_{\delta_1},\cdots,G_{\delta_n},\cdots$ 为总体 $G_\delta(R)$ 的一个序贯样本. 对于 $x\in S$, 命 $D_N(x)$ 表示前 $N$ 个随机带中覆盖 $x$ 的带子个数. 若 $S$ 上有一点被 $m$ 个带子覆盖, 此处 $m$ 是一个给定的正整数, 我们就称这个球形轧滚作废了. 对于正整数 $m$, 命 $T_m$ 表示对某 $\boldsymbol{x}\in S$ 满足 $D_N(\boldsymbol{x})\geqslant m$ 的最小 $N$, 即

$$T_m=\min\{N: D_N(\boldsymbol{x})\geqslant m \text{ 对于某个 } \boldsymbol{x}\in S\}. \tag{4.3}$$

则 $T_m$ 就是球形轧滚的寿命. 我们希望求得 $T_m$ 的分布, 从而找到延长球形轧滚的寿命的途径.

给出 $T_m$ 的分布函数的表达式是困难的, 所以我们用模拟处理这一问题. 我们用到下面的事实:

由 [13] 例 3, 我们有 $S$ 上均匀散布的点集 $\{\boldsymbol{x}_k=(x_{k1},x_{k2},x_{k3})', k=1,\cdots,n\}$. 详细之

$$\begin{cases} x_{k1}=\cos(\pi c_{k1}), \\ x_{k2}=\sin(\pi c_{k1})\cos(2\pi c_{k2}), k=1,\cdots,n \\ x_{k3}=\sin(\pi c_{k1})\sin(2\pi c_{k2}), \end{cases} \tag{4.4}$$

此处 $F_i(c_{ki})=b_{ki}, i=1,2, k=1,\cdots,n, \{\boldsymbol{b}_k=(b_{k1},b_{k2})'k=1,\cdots,n\}$ 为一个在 $I^2$ 上均匀散布的点集，及

$$F_i(x)=\frac{\pi}{B\left(\frac{1}{2},\frac{3-i}{2}\right)}\int_0^x(\sin\pi t)^{(3-i-1)}\mathrm{d}t, \quad i=1,2,$$

即

$$F_1(x)=\frac{1}{2}(1-\cos(\pi x)),$$
$$F_2(x)=x.$$

所以

$$b_{k1} = \frac{1}{2}(1 - \cos(\pi c_{k1})),$$
$$b_{k2} = c_{k2},$$

及

$$\begin{cases} x_{k1} = 1 - 2b_{k1}, \\ x_{k2} = 2\sqrt{b_{k1} - b_{k1}^2}\cos(2\pi b_{k2}), \\ x_{k3} = 2\sqrt{b_{k1} - b_{k1}^2}\sin(2\pi b_{k2}), \quad k = 1, 2, \cdots, n. \end{cases} \tag{4.5}$$

命

$$T_m^* = \min\{N : D_N(\boldsymbol{x}_k) \geqslant m, \text{对于某个 } k, 1 \leqslant k \leqslant n\}.$$

当 $n$ 充分大时, $T_m^*$ 接近于 $T_m$ 及 $T_m^*$ 的分布接近于 $T_m$ 的分布. 我们基于对 $T_m^*$ 进行模拟.

对于 $S$ 上一点 $\boldsymbol{v}$, 它对应于一个大圆 $R$ 使包含 $R$ 的平面的法线正好就是 $\overrightarrow{\boldsymbol{ov}}$. 若将 $\boldsymbol{v}$ 与 $-\boldsymbol{v}$ 等同起来, 则 $S$ 上的点与大圆间有一个一一对应, 从而与 $S$ 上以 $\delta$ 为宽度的带子之间亦有一一对应. 因此生成在 $S$ 上均匀散布的带子 $G_\delta(R)$ 与在 $S$ 生成一个均匀散布点集 $\boldsymbol{v}$ 是等价的. 我们也用 $G_\delta(\boldsymbol{v})$ 表示对应于 $\boldsymbol{v}$ 的带子. 我们的模拟包括以下步骤:

第 1 步, 给予 $m$ 与 $\delta$, 例如 $m = 20$, $\delta = 0.2$.

第 2 步, 选取一个适当的 $n$ (例如 $n = 1069$) 并生成一个在 $S$ 上均匀散布的点集 $\{\boldsymbol{x}_k, k = 1, \cdots, n\}$.

第 3 步, 由模拟的标准技术, 序贯地生成在 $S$ 上均匀散布的点 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots$ 从而得到对应的带子 $G_\delta(\boldsymbol{v}_1), G_\delta(\boldsymbol{v}_2), \cdots$

第 4 步, 若有 $N(N = 1, 2, \cdots)$ 个随机带子生成了, 记 $D_N(\boldsymbol{x}_k)$ 为这 $N$ 个带子中覆盖 $\boldsymbol{x}_k$ 者的个数. 若对于某个 $k$ 有 $D_N(\boldsymbol{x}_k) = m$, 则转入第 5 步, 否则就回到第 3 步并生成第 $N+1$ 个随机带子.

第 5 步, 计算已经生成的带子数目, 这就是 $T_m^*$ 的一个观测值.

将上面步骤重复 $n_0$ 次, 我们就得到一个大小为 $n_0$ 的 $T_m^*$ 的样本. 取 $n_0 = 5000$, 对应的样本均值与样本标准离差为

$$\overline{T}_m^* = 99.7 \text{ 及 } \sigma(T_m^*) = 9.8.$$

进而言之, 对应的经验分布接近于正态分布.

用同样的方法, 我们得到大小为 $n_0 = 5000$ 的 20 个样本 (总共 100 000 个观测值), 发现它们的结果彼此很相近.

由于 $T_m$(或 $T_m^*$) 表示轧滚的寿命, 我们注意到上述模拟的 100 000 次观测中, $T_m^*$ 可达 125. 我们将它们对应的法方向记为 $\boldsymbol{v}_1^*, \boldsymbol{v}_2^*, \cdots, \boldsymbol{v}_{125}^*$. 这表示在 $\delta = 0.2$

与 $m=20$ 的情况下, 若 $\boldsymbol{v}_i=\boldsymbol{v}_i^*, i=1,2,\cdots,125$, 我们总能得到 $T_m^*=125$, 这比上述随机选取 $\{\boldsymbol{v}_i\}$ 好些. 是否可能找到比 $\{v_i^*, i=1,2,\cdots,125\}$ 更好的另一集合 $\boldsymbol{v}_1^{**},\boldsymbol{v}_2^{**},\cdots$ 呢? 我们用 $I^2$ 上的一个均匀散布集合去产生 $S$ 上的集合 $\{\boldsymbol{v}_k^{**}, k=1,\cdots,n\}$. 首先我们取 $n=126$ 并发现 $T_m^*=126$. 然后一个个地增加 $n$ 直至 $T_n^*$ 不能增加为止, 最后得到一个集合 $\{v_k^{**}=(v_{k1},v_{k2},v_{k3})', k=1,2,\cdots,155\}$, 由此得到 $T_m^*=155$! 其中 $v_k^{**}$ 由

$$\begin{cases} v_{k1}=1-2b_{k1}, \\ v_{k2}=2\sqrt{b_{k1}-b_{k1}^2}\cos(2\pi b_{k2}), \\ v_{k3}=2\sqrt{b_{k1}-b_{k2}^2}\sin(2\pi b_{k2}), \end{cases}$$

给出, 此处 $\{\boldsymbol{b}_k=(b_{k1},b_{k2})', k=1\cdots,155\}$ 是由 $(h_1,h_2;n)=(1,20,155)$ 产生的 (见 [13], §3).

这说明在我们模拟中, 数论方法取胜于蒙特卡罗方法 100 000 次实验.

感谢: 我们感谢魏刚先生很卓越的计算机工作.

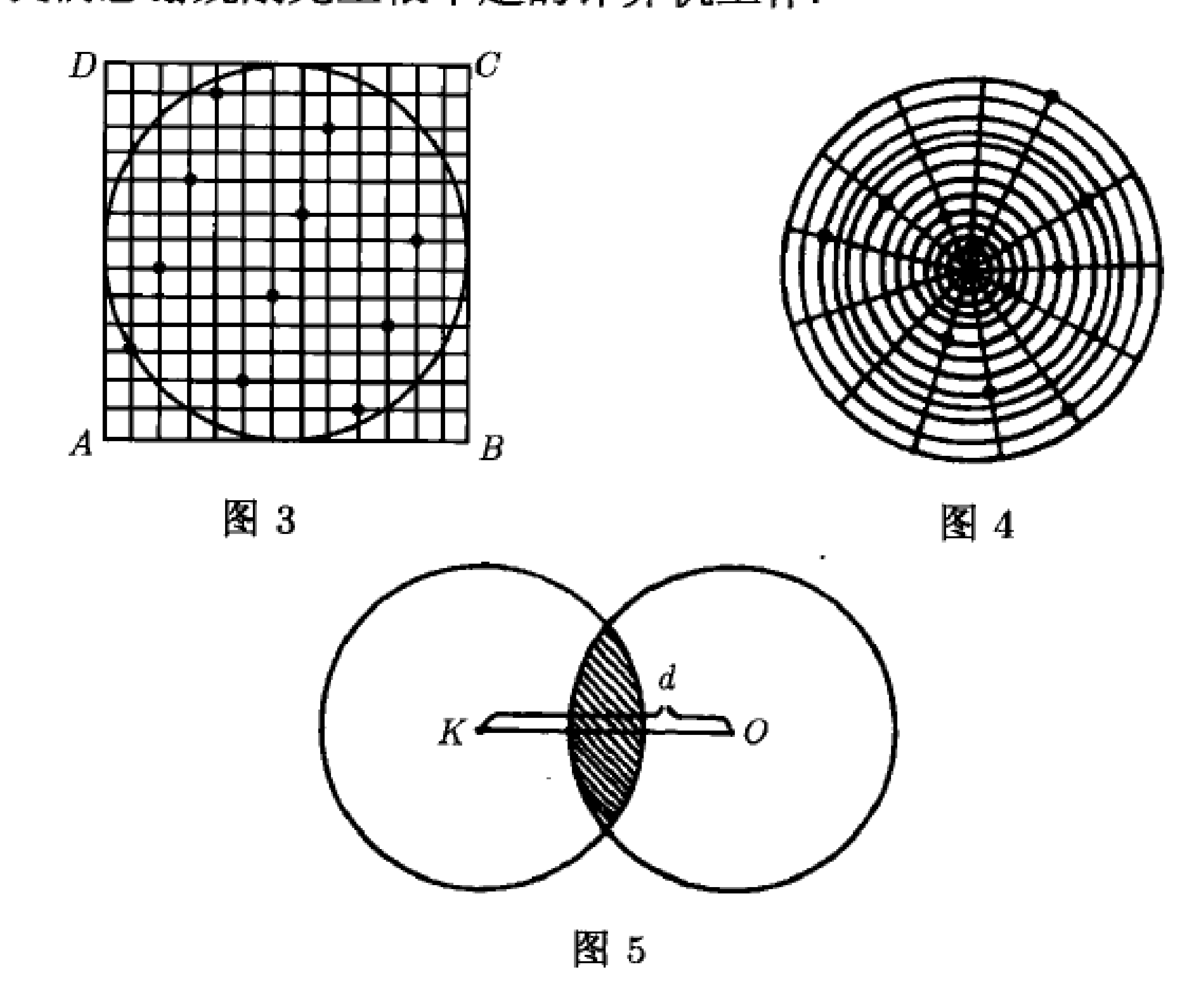



图 3　　图 4

图 5

## 参 考 文 献

[1] 成平 (Cheng, P.), An open problem in steel rolling, *Mathematics in Practice and Theory*, **2** (1983), 79～79.

[2] Cornell, J. A., Experiments with mixtures: A review. *Technometrics*, **15** (1973), 437～455.

[3] Cornell, J. A., Some comments on designs for Cox's mixture polynomial, *Technometrics*, **17** (1975). 25~35.

[4] Cornell, j. A., Exporiments with Mixtures, designs, models, and the analysis of mixture data, 1981, Wiley. New York.

[5] Cox, D. R., A note on polynomial response functions for mixtures, *Biometrika*, **58** (1971), 155~159.

[6] Draper, N. R., & Lawrence, W. E., Mixture designs for three factors, *J. Royal Statist. Soc. B*, **27** (1965), 450~465.

[7] 华罗庚与王元 (Hua, L. K.,&Wang, Y.), *Applications of Number Theory to Numerial Analysis,* Springer-Verlag and Science Press, 1981, Berlin/Beijing.

[8] Scheffe, H., Experiments with mixtures, *J. Royal Statist, Soc. B*, **20** (1958), 344~360.

[9] Scheffe, H., The simplex-centroid design for experiments with mixtures, *J. Royal Statist. Soc. B*, **25** (1963), 235~263.

[10] Snee, R. D., Techniques for the analysis of mixture data, *Technometrics*, **15** (1973), 517~528.

[11] Thompson, W. O. & Myers, R. H., Response surface design for experimonts with mixtures, *Technometrics*, **10** (1968), 739~756.

[12] 王元与方开泰 (Wang, Y. & Fang, K. T.), Anote on uniform distribution and experimental design, *Kexue Yongba*, **26** (1981), 485~489.

[13] 王元与方开泰 (Wang, Y. & Fang, K. T.), Number theoretic methods in applied statistics, *Chin. Ann of Math.*, **11B**: 1(1990), 41~55.

# 混料均匀设计 *

王元
(中国科学院数学研究所)

方开泰
(香港浸会大学, 中国科学院应用数学研究所)

## 摘　　要

考虑混料试验设计: $0 \leqslant a_i < x_i < b_i \leqslant 1, 1 \leqslant i \leqslant s, x_1 + \cdots + x_s = 1$, 此处 $a_i, b_i, 1 \leqslant i \leqslant s$ 为给予常数. 用数论中的一致分布理论对这一模型提供一个处理方法.

**关键词　试验设计　均匀设计　混料均匀设计　数论网 (NT- 网)　偏差**

作为数论中均匀分布理论对试验设计的应用, 建议用一个所谓的均匀设计法 [1]. 假定有 $s$ 个因素 $x_1, \cdots, x_s$, 不失一般性, 假定试验区域为单位立方体 $C^s = [0,1]^s$ 及 $C^s$ 中每一点皆对应于一个实验, 均匀设计的思想为在 $C^s$ 中找一个 $n$ 个点的集合.

$$\mathscr{F} = \{\boldsymbol{c}_k = (c_{k_1}, \cdots, c_{k_s}), 1 \leqslant k \leqslant n\}.$$

它在 Weyl[2] 的意义之下有低的偏差. 若 $D(\mathscr{F}) = o(n^{-\frac{1}{2}})$ (当 $n \to \infty$), 则称 $\mathscr{F}$ 为一个数论网或 NT- 网.

一系列方法可以产生 NT- 网, 对于不同的 $n$ 与 $s$ 已编成表, 见文献 [3, 4]. 然后可以在 $\mathscr{F}$ 上安排 $n$ 次试验, 并得到 $\mathscr{F}$ 中的一个点 $\boldsymbol{c}$. 在这一点的试验结果是 $\mathscr{F}$ 上 $n$ 个试验结果的最佳者, 但在实际应用时, 特别在化学试验与化学工程中, 因素之间需加上一些约束条件. 例如

$$0 \leqslant a_i < x_i < b_i \leqslant 1, 1 \leqslant i \leqslant s, \sum_{i=1}^{s} x_i = 1. \tag{1}$$

此处 $a_i, b_i$ 为给予的常数. 我们称有约束条件的试验设计为混料试验设计. Cornell[5] 对混料试验设计给出一个全面的阐述. 例如, 他将区域 (1) 换成一个椭球, 然后将原

* 原载《中国科学》(A 辑) 第 26 卷第 1 期, 1996 年, 1~10.

** 国家自然科学基金和数学研究所 (台北) 资助项目.

来的问题变为一个 $(s-1)$- 维的问题. 本文的宗旨在于将均匀分布理论用于带约束条件 (1) 的试验设计中去. 模型 (1) 的最简单的情况为 $a_i = 0$ 与 $b_i = 1, 1 \leqslant i \leqslant s$. 这一问题等价于在单纯形

$$S = \left\{ \boldsymbol{x} = (x_1, \cdots, x_s) : x_i \geqslant 0, 1 \leqslant i \leqslant s, \sum_{i=1}^{s} x_i = 1 \right\}$$

上找出一个均匀散布的集合 $\{\boldsymbol{x}_k, 1 \leqslant k \leqslant n\}$. 我们曾建议过一个由 $C^{s-1}$ 上 $n$ 个点的数论 - 网导出 $S$ 上均匀散布的 $n$ 个点的方法 [4,6,7]. 我们将推广我们的方法来处理更为困难与一般的情况 (1), 即给出寻找区域 (1) 上均匀散布的集合的方法.

## §1. 逆变换方法

命 $D$ 为 $s$- 维欧氏空间 $\boldsymbol{R}^s$ 中的一个有界区域及 $\boldsymbol{x} = (x_1, \cdots, x_s)$ 为 $D$ 上定义的, 且有积垒分布函数 (c.d.f)$\boldsymbol{F}(\boldsymbol{X})$ 的一个随机矢量, $\boldsymbol{x}$ 有随机表示:

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{\varphi}), \boldsymbol{\varphi} \in C^t,$$

此处 $t \leqslant s$ 及 $\boldsymbol{\varphi} = (\varphi_1, \cdots, \varphi_t) \in C^t$ 为有独立分量的随机矢量. 假定 $\dfrac{\partial x_j}{\partial \varphi_i}, 1 \leqslant i \leqslant t$, $1 \leqslant j \leqslant s$ 在 $C^t$ 上是连续的. 记

$$\boldsymbol{T} = \left( \frac{\partial x_j}{\partial \varphi_i} \right), 1 \leqslant i \leqslant t, 1 \leqslant j \leqslant s$$

及

$$J(\boldsymbol{\varphi}) = \det(\boldsymbol{T}\boldsymbol{T}')^{1/2},$$

$J(\boldsymbol{\varphi})$ 是 $D$ 关于 $\boldsymbol{x}$ 的体积元素. 当 $t = s, J(\boldsymbol{\varphi})$ 就是通常的 Jacobian. 由 $\varphi_i's$ 的独立性可知

$$\frac{1}{V(D)} J(\boldsymbol{\varphi}) = \prod_{i=1}^{s} p_i(\varphi_i),$$

此处 $V(D)$ 为 $D$ 的体积及 $p_i(\varphi)$ 为 $\varphi_i$ 的概率密度函数 (p.d.f.), $1 \leqslant i \leqslant t$. 所以

$$\frac{1}{V(D)} \int_{\boldsymbol{\varphi} \leqslant \boldsymbol{r}} J(\boldsymbol{\varphi}) \mathrm{d}\boldsymbol{\varphi} = \prod_{i=1}^{t} F_i(r_i), \tag{2}$$

此处 $\mathrm{d}\boldsymbol{\varphi} = \prod_{i=1}^{t} \mathrm{d}\varphi_i, \boldsymbol{\varphi} \leqslant \boldsymbol{r} = (r_1, \cdots, r_t)$ 表示 $\varphi_i \leqslant r_i, 1 \leqslant i \leqslant t, F_i(r)$ 表示 $\varphi_i$ 的 c.d.f. $1 \leqslant i \leqslant t$.

命 $\mathscr{F}=\{\boldsymbol{x}_k, 1\leqslant k\leqslant n\}$ 为 $D$ 上的一个点集及 $x_i=\boldsymbol{x}(\boldsymbol{\varphi}_i), 1\leqslant i\leqslant n$. 命

$$N(\mathscr{F},\boldsymbol{r})=\sum_{i=1}^{n} I(\boldsymbol{\varphi}_i\leqslant \boldsymbol{r}),$$

其中 $I(A)$ 表示 $A$ 的指标函数, 即

$$I(A)=\begin{cases}1, \text{当 } A \text{ 出现}, \\ 0, \text{其他情况}.\end{cases}$$

则 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 的经验分布函数可以定义为

$$F_n(\boldsymbol{r})=\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{\varphi}_i\leqslant \boldsymbol{r}).$$

于是

$$\sup_{r\in C^t}|F_n(\boldsymbol{r})-F(\boldsymbol{r})| \tag{3}$$

为 $F(\boldsymbol{r})$ 关于拟合最优检验中的 Kolmogorov-Smirnov 距离. 若 $F(\boldsymbol{r})$ 为 $C^t$ 上的均匀分布, 则由 (2) 式可知

$$F(\boldsymbol{r})=\frac{V(\boldsymbol{t}\leqslant \boldsymbol{r})}{V(D)}=\prod_{i=1}^{t} F_i(r_i),$$

此处

$$V(\boldsymbol{t}\leqslant \boldsymbol{r})=\int_{t\leqslant r} J(\boldsymbol{\varphi})\mathrm{d}\boldsymbol{\varphi}.$$

将 (3) 式记为

$$D_F(\mathscr{F})=\sup_{r\in C^t}\left|F_n(\boldsymbol{r})-\frac{V(\boldsymbol{t}\leqslant \boldsymbol{r})}{V(D)}\right|.$$

这是集合 $\mathscr{F}$ 在 $D$ 上均匀度的一个测度, 它称为 $\mathscr{F}$ 的 $F$- 偏差. 当 $s=t, D=C^t$ 及 $x_i=\varphi_i, 1\leqslant i\leqslant s, D_F(\mathscr{F})$ 就是 $\mathscr{F}$ 在 Weyl[2] 意义下的偏差, 并记为 $D(\mathscr{F})$. 命

$$\mathscr{F}=\{\boldsymbol{c}_k=(c_{k1},\cdots,c_{kt}), 1\leqslant k\leqslant n\}$$

为 $C^t$ 上均匀分布的一个样本, 它有偏差 $D(\mathscr{F})$. 我们用 $\boldsymbol{F}_i^{-1}(\boldsymbol{r})$ 表示 $\boldsymbol{F}_i(\boldsymbol{r})$ 的逆函数, $1\leqslant i\leqslant t$, 及

$$\boldsymbol{x}_k=\boldsymbol{x}(\boldsymbol{F}^{-1}(\boldsymbol{c}_k)),$$

此处

$$\boldsymbol{F}^{-1}(\boldsymbol{c}_k)=(\boldsymbol{F}_1^{-1}(c_{k1}),\cdots,F_t^{-1}(c_{kt})), 1\leqslant k\leqslant n.$$

今往寻求集合 $\mathscr{F}^* = \{\boldsymbol{x}_k = \boldsymbol{x}(\boldsymbol{F}^{-1}(\boldsymbol{c}_k)), 1 \leqslant k \leqslant n\}$ 的 $F$- 偏差. 由于

$$F_n(\boldsymbol{r}) = \frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{F}^{-1}(\boldsymbol{c}_k) \leqslant \boldsymbol{r}) = \frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{c}_k \leqslant \boldsymbol{F}(\boldsymbol{r})),$$

所以

$$\begin{aligned} D_F(\mathscr{F}^*) &= \sup_{r\in C^t}\left|\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{c}_k \leqslant F(\boldsymbol{r})) - \prod_{i=1}^{t} F_i(r_i)\right| \\ &= \sup_{u\leqslant C^t}\left|\frac{1}{n}\sum_{i=1}^{n} I(\boldsymbol{c}_k \leqslant \boldsymbol{u}) - \prod_{i=1}^{t} u_i\right| = D(\mathscr{F}), \end{aligned}$$

这表示 $\mathscr{F}^*$ 的 $F$- 偏差正好是 $D(\mathscr{F})$. 因此从 $C^t$ 上一个有偏差 $d$ 的集合出发, 就可以导出 $D$ 上一个有 $F$- 偏差 $d$ 的集合 $\mathscr{F}^*$. 这一方法称为逆变换法, 我们常常称有 $F$- 偏差 $o(n^{-\frac{1}{2}})$ 的 $n$ 个点的集合为 NT- 网. 我们可以构造具有 $F$- 偏差 $O(n^{-1}(\log n)^{s-1})$ 的 NT- 网 [3,4].

## §2. 区域 $S(a,b)$

假定 $\boldsymbol{a} = (a_1, \cdots, a_s)$ 与 $\boldsymbol{b} = (b_1, \cdots, b_s)$ 为两个适合

$$0 \leqslant a_i < b_i \leqslant 1, 1 \leqslant i \leqslant s$$

的矢量, 及

$$a = \sum_{i=1}^{s} a_i < 1 \text{ 与 } b = \sum_{i=1}^{s} b_i > 1. \tag{4}$$

记 $S(\boldsymbol{a},\boldsymbol{b})$ 为区域

$$S(\boldsymbol{a},\boldsymbol{b}) = \left\{\boldsymbol{x} = (x_1, \cdots, x_s) : a_i < x_i < b_i, 1 \leqslant i \leqslant s, \sum_{i=1}^{s} x_i = 1\right\}.$$

注意 (4) 式是 $S(\boldsymbol{a},\boldsymbol{b})$ 非空的充分且必要的条件, 事实上若 (4) 式成立, 则显然 $S(\boldsymbol{a},\boldsymbol{b})$ 非空. 反之, 若 $a \geqslant 1$ 或 $b \leqslant 1$, 则对于所有适合 $a_i < x_i < b_i, 1 \leqslant i \leqslant s$ 的 $\boldsymbol{x}$ 皆有 $\sum\limits_{i=1}^{s} x_i > 1$ 或 $\sum\limits_{i=1}^{s} x_i < 1$. 命

$$y_i = \frac{x_i - a_i}{b_i - a_i}, 1 \leqslant i \leqslant s, \tag{5}$$

则

$$\sum_{i=1}^{s}(b_i - a_i)y_i = \sum_{i=1}^{s} x_i - \sum_{i=1}^{s} a_i = 1 - a.$$

记

$$d_i = \frac{b_i - a_i}{1 - a}, 1 \leqslant i \leqslant s,$$

则得 $\sum_{i=1}^{s} d_i y_i = 1, d_i > 0, 1 \leqslant i \leqslant s$. 因此 $S(\boldsymbol{a}, \boldsymbol{b})$ 与条件 (4) 分别变成

$$S(\boldsymbol{d}) = \left\{ \boldsymbol{y} = (y_1, \cdots, y_s) : y \in C^s, \sum_{i=1}^{s} d_i y_i = 1, d_i > 0, 1 \leqslant i \leqslant s \right\}$$

与

$$\sum_{i=1}^{s} d_i > 1. \tag{6}$$

其中 $\boldsymbol{d} = (d_1, \cdots, d_s)$. 命

$$S^*(\boldsymbol{d}) = \left\{ \boldsymbol{y} = (y_1, \cdots, y_s) : y_i > 0, 1 \leqslant i \leqslant s, \sum_{i=1}^{s} d_i y_i = 1, d_i > 0, 1 \leqslant i \leqslant s \right\}.$$

首先给出在 $S^*(\boldsymbol{d})$ 找出一个 NT- 网的方法. 变换

$$\begin{cases} d_j y_j = z_1 \cdots z_{j-1}\ (1 - z_j),\ 1 \leqslant j \leqslant s - 1, \\ d_s y_s = z_1 \cdots z_{s-1}, \boldsymbol{z} = (z_1, \cdots, z_{s-1}) \in C^{s-1} \end{cases}$$

为 $C^{s-1}$ 与 $S^*(\boldsymbol{d})$ 间的一个一一对应. 命

$$\Delta = (\det \boldsymbol{H}\boldsymbol{H}')^{1/2} \text{ 及 } \boldsymbol{H} = \left( \frac{\partial y_j}{\partial z_i} \right), 1 \leqslant i \leqslant s - 1, 1 \leqslant j \leqslant s, \tag{7}$$

则

$$\Delta = C(\boldsymbol{d}) z_1^{s-2} \cdots z_{s-2}, \tag{8}$$

其中

$$C(\boldsymbol{d}) = (d_1 \cdots d_s)^{-1} \left( \sum_{i=1}^{s} d_i^2 \right)^{1/2}.$$

证明将在第 4 节中给出. 因此 $S^*(\boldsymbol{d})$ 的体积等于

$$\begin{aligned} V(S^* \boldsymbol{d}) &= \int_{C^{s-1}} \Delta \mathrm{d}z = C(\boldsymbol{d}) \int_{C^{s-1}} z_1^{s-2} \cdots z_{s-2} \mathrm{d}z_1 \cdots \mathrm{d}z_{s-1} \\ &= C(\boldsymbol{d})/(s-1)!. \end{aligned}$$

由此可见 $z_1, \cdots, z_{s-1}$ 是相互独立的, $z_j$ 的 p.d.f. 与 c.d.f. 分别是

$$p_j = (s - j) z^{s-j-1}$$

与

$$G_j = \int_0^z p_j(t)\mathrm{d}t = z^{s-j}, \quad z \in C^1, \quad 1 \leqslant j \leqslant s-1.$$

我们可以用逆变换法得到 $S^*(\boldsymbol{d})$ 上一个 NT- 网：由 $C^{s-1}$ 上的一个 NT- 网

$$\{\boldsymbol{c}_k = (c_{kl}, \cdots c_{k,s-1}), 1 \leqslant k \leqslant n\}$$

出发 (见文献 [4] 附录), 由于 $G_j(t)$ 的逆函数为

$$G_j^{-1}(z) = z^{\frac{1}{s-j}}, 1 \leqslant j \leqslant s-1,$$

故得 $S^*(\boldsymbol{d})$ 上一个 NT- 网

$$\{\boldsymbol{y}_k = (y_{k1}, \cdots, y_{ks}), 1 \leqslant k \leqslant n\}, \tag{9}$$

其中

$$\begin{cases} y_{kj} = d_j^{-1} c_{k1}^{\frac{1}{s-1}} \cdots c_{k,j-1}^{\frac{1}{s-j+1}} (1 - c_{kj}^{\frac{1}{s-j}}), \\ y_{kj} = d_s^{-1} c_{k1}^{\frac{1}{s-1}} \cdots c_{k,s-2}^{\frac{1}{2}} c_{k,s-1}, 1 \leqslant j \leqslant s-1, 1 \leqslant k \leqslant n. \end{cases}$$

代入 (5) 式, 则得区域

$$S(\boldsymbol{a}) = \left\{ \boldsymbol{x} = (x_1, \cdots, x_s) : 1 > x_i > a_i, 1 \leqslant i \leqslant s, \sum_{i=1}^s x_i = 1 \right\}$$

上一个 NT- 网. 此处

$$\begin{cases} x_{kj} = (1-a) c_{k1}^{\frac{1}{s-1}} \cdots c_{k,j-1}^{\frac{1}{s-j+1}} (1 - c_{kj}^{\frac{1}{s-j}}) + a_j, \\ x_{ks} = (1-a) c_{k1}^{\frac{1}{s-1}} \cdots c_{k,s-2}^{\frac{1}{2}} c_{k,s-1} + a_s, 1 \leqslant j \leqslant s-1, 1 \leqslant k \leqslant n. \end{cases}$$

由于变换 (5) 的 Jacobian 为常数 $\prod_{i=1}^s (b_i - a_i)$, 所以 $\{\boldsymbol{x}_k, 1 \leqslant k \leqslant n\}$ 与 $\{\boldsymbol{y}_k : 1 \leqslant k \leqslant m\}$ 的 $F$- 偏差有相同阶. 于是在以下仅考虑 $S(\boldsymbol{d})$.

## §3. $S^*(d)$ 与 $S(d)$ 的体积

我们的目的为在 $S(\boldsymbol{d})$ 上得到一个 $n$ 个点的 NT- 网. 先在 $S^*(\boldsymbol{d})$ 上选取一个有 $m$ 个点 NT- 网 (9) 式, 它有差不多 $n$ 个点落在 $S(\boldsymbol{d})$ 上, 因 $\mathscr{F}$ 在 $S^*(\boldsymbol{d})$ 上均匀散布, 所以 $\frac{m}{n}$ 应渐近地等于 $S^*(\boldsymbol{d})$ 与 $S(\boldsymbol{d})$ 的体积之比, 即

$$r \equiv \frac{m}{n} \approx \frac{V(S^*(\boldsymbol{d}))}{V(S(\boldsymbol{d}))}.$$

因此 $m$ 的估计即归结为 $V(S^*(\boldsymbol{d}))$ 与 $V(S(\boldsymbol{d}))$ 之估计. 由 (7) 式可知

$$V(S^*(\boldsymbol{d}))=C(\boldsymbol{d})\int_{z\in C^{s-1}} z_1^{s-2}\cdots z_{s-2}\mathrm{d}z_1\cdots\mathrm{d}z_{s-1}=C(\boldsymbol{d})/(s-1)!. \tag{10}$$

今估计 $V(S(\boldsymbol{d}))$,

$$V(S(\boldsymbol{d}))=C(\boldsymbol{d})N(\boldsymbol{d}), \tag{11}$$

其中

$$N(\mathrm{d})=\int_{\substack{z\in C^{s-1}\\ y\in C^{s}}} z_1^{s-2}\cdots z_{s-2}\mathrm{d}z_1\cdots\mathrm{d}z_{s-1}.$$

我们第一步找出一个递推公式, 它将 $(s-1)$- 维积分 $N(\boldsymbol{d})$ 归结为一个 $(s-2)$- 维积分. 第二步求出 $s=2$ 时, $N(\boldsymbol{d})$ 的表达式. 最后逐次运用递推公式, 求出 $s=3,4,\cdots$ 时, $N(\boldsymbol{d})$ 的表达式.

(1) 约化. 不失一般性, 我们可以假定

$$d_1\geqslant d_2\geqslant\cdots\geqslant d_s. \tag{12}$$

否则, 我们可以改变变数的次序使 (12) 式成立.

1) 若 $d_s\geqslant 1$, 则 $S(\boldsymbol{d})=S^*(d)$, 及

$$N(\boldsymbol{d})=\frac{1}{(s-1)!}. \tag{13}$$

2) 若 $d_1\geqslant 1$ 及 $d_s<1$, 则 $z_1$ 没有约束, 即 $z_1$ 可以取 $C^1$ 之任何值, 关于给予 $z_1\in C^1, z_1\neq 0$, 考虑区域:

$$S^*\left(\frac{d_2}{z_1},\cdots\frac{d_s}{z_1}\right):\begin{cases}\dfrac{d_i}{z_1}y_i=z_2\cdots z_{i-1}(1-z_i),\\ \dfrac{d_s}{z_1}y_s=z_2\cdots z_{s-1}, 2\leqslant i\leqslant s-1.\end{cases}$$

此处 $(z_2,\cdots,z_{s-1})\in C^{s-2},(y_2,\cdots,y_s)\in C^{s-1}$. 当 $z_1\in(0,d_s)$ 时, $\dfrac{d_s}{z_1}\geqslant 1$ 及在 (13) 式中将 $s$ 换为 $s-1$, 得

$$\begin{aligned}N(\boldsymbol{d})=&\int_0^{d_s} z_1^{s-2}\mathrm{d}z_1\int_{S^*(\frac{d_2}{z_1},\cdots,\frac{d_s}{z_1})} z_2^{s-3}\cdots z_{s-2}\mathrm{d}z_2\cdots\mathrm{d}z_{s-1}\\ &+\int_{d_s}^1 z_1^{s-2}\mathrm{d}z_1\int_{S^*(\frac{d_2}{z_1},\cdots,\frac{d_s}{z_1})} z_2^{s-3}\cdots z_{s-2}\mathrm{d}z_2\cdots dz_{s-1}\\ =&\frac{1}{(s-2)!}\int_0^{d_s} z_1^{s-2}\mathrm{d}z_1+\int_{d_s}^1 z_1^{s-2}\mathrm{d}z_1\int_{S^*(\frac{d_2}{z_1},\cdots,\frac{d_s}{z_1})} z_2^{s-3}\cdots z_{s-2}\mathrm{d}z_2\cdots\mathrm{d}z_{s-1}.\end{aligned}$$

由 (4)、(6) 式可知, $z_1$ 必须满足

$$\sum_{i=2}^{s} d_i \geqslant z_1,$$

否则后一积分值为 0, 因此得递推公式

$$N(\boldsymbol{d}) = \frac{d_s^{s-1}}{(s-1)!} + \int_{d_s}^{\min(1,\sum_{u=2}^{s} d_i)} z_1^{s-2} N\left(\frac{d_2}{z_1}, \cdots, \frac{d_s}{z_1}\right) \mathrm{d}z_1. \tag{14}$$

3) 若 $d_1 < 1$, 则 $z_1$ 必须满足 $1 - z_1 \leqslant d_1$, 这是由于 $d_1 y_1 = 1 - z_1$ 及 $y_1 \in C^1$ 之故. 所以得递推公式

$$N(\boldsymbol{d}) = \int_{1-d_1}^{\min(1,\sum_{i=2}^{s} d_i)} z_1^{s-2} N\left(\frac{d_2}{z_1}, \cdots, \frac{d_s}{z_1}\right) \mathrm{d}z_1. \tag{15}$$

(2) $s = 2$. $N(\boldsymbol{d})$ 等于适合于

$$d_1 y_1 = 1 - z_1 d_2 y_2 = z_1, z_1 \in C^1$$

的 $z_1$ 的测度. 所以

$$N(\boldsymbol{d}) = d_2 (d_1 \geqslant 1, d_2 < 1), \tag{16}$$

$$N(\boldsymbol{d}) = |1 - d_1 < z_1 < d_2| = d_1 + d_2 - 1 (d_1 < 1). \tag{17}$$

(3) $s = 3$. 若 $d_2 \geqslant 1, d_3 < 1$, 则由 (14) 与 (16) 式得

$$N(\boldsymbol{d}) = \frac{d_3^2}{2} + \int_{d_3}^{\min(1,d_2+d_3)} z_1 N\left(\frac{d_2}{z_1}, \frac{d_3}{z_1}\right) \mathrm{d}z_1 = \frac{d_3^2}{2} + \int_{d_3}^{1} z_1 \frac{d_3}{z_1} \mathrm{d}z_1 = d_3 - \frac{d_3^2}{2}. \tag{18}$$

类似地, 由 (13) ~ (17) 式可得

$$\begin{aligned} N(\boldsymbol{d}) = &(d_2 + d_3) \min(1, d_2 + d_3) \\ &- \frac{\min(1, d_2 + d_3)^2}{2} - \frac{d_2^2 + d_3^2}{2} \quad (d_1 \geqslant 1, d_2 < 1), \end{aligned} \tag{19}$$

$$\begin{aligned} N(\boldsymbol{d}) = &d_1 - \frac{1}{2}(d_1^2 + d_2^2 + d_3^2 + 1) + (d_2 + d_3) \min(1, d_2 + d_3) \\ &- \frac{\min(1, d_2 + d_3)^2}{2} \quad (d_1 < 1, 1 - d_1 < d_3), \end{aligned} \tag{20}$$

$$N(\boldsymbol{d}) = d_3 \left(d_1 + d_2 - 1 + \frac{1}{2} d_3\right) \quad (d_1 < 1, d_3 \leqslant 1 - d_1 < d_2), \tag{21}$$

$$N(\boldsymbol{d})=\frac{1}{2}(d_1+d_2+d_3-1)^2 \quad (d_1<1, d_2\leqslant 1-d_1). \tag{22}$$

当 $s\geqslant 4$, 由于 $N(\boldsymbol{d})$ 的解析表达式太复杂, 所以建议用统计模拟方法来估计比值 $r$: 在 $C^{s-1}$ 上取一个 NT- 网

$$\mathscr{F}=\{\boldsymbol{c}_k=(c_{k1},\cdots,c_{k,s-1}),1\leqslant k\leqslant m\},$$

此处 $m$ 充分大. 若 $\mathscr{F}$ 有偏差 $d$, 则由它诱导的 $S^*(\boldsymbol{d})$ 上的集合 $\mathscr{F}^*=\{\boldsymbol{y}_k=(y_{k1},\cdots,y_{ks}),1\leqslant k\leqslant n\}$ 有 $F$- 偏差 $d$ (见 (9) 式). 由于 $\mathscr{F}^*$ 在 $S^*(\boldsymbol{d})$ 上均匀散布, 所以 $\mathscr{F}^*$ 落于 $S(\boldsymbol{d})$ 上的点数渐近地等于

$$\left[m\frac{V(S(\boldsymbol{d}))}{V(S^*(\boldsymbol{d}))}\right]=n,$$

其中 $[x]$ 表示 $x$ 的整数部分, 即 $\dfrac{m}{n}$ 可以看作是 $\dfrac{V(S^*(\boldsymbol{d}))}{V(S(\boldsymbol{d}))}$ 的近似值.

## §4. $\Delta$ 的计算

回忆 (7) 与 (8) 式, $\Delta=(\det \boldsymbol{H}\boldsymbol{H}')^{1/2}$, 此处

$$\boldsymbol{H}=\left(\frac{\partial y_j}{\partial z_i}\right)=$$
$$\begin{pmatrix} -d_1^{-1}, & d_2^{-1}(1-z_2), & \cdots, & d_{s-1}^{-1}\dfrac{z_1\cdots z_{s-2}}{z_1}(1-z_{s-1}), & d_s^{-1}\dfrac{z_1\cdots z_{s-1}}{z_1} \\ 0, & -d_2^{-1}z_1, & \cdots, & d_{s-1}^{-1}\dfrac{x_1\cdots z_{s-2}}{z_2}(1-z_{s-1}), & d_s^{-1}\dfrac{x_1\cdots z_{s-1}}{z_2} \\ \vdots & \vdots & \ddots & & \\ 0, & 0, & \cdots, & -d_{s-1}^{-1}\dfrac{z_1\cdots z_{s-1}}{z_{s-1}}, & d_s^{-1}\dfrac{z_1\cdots z_{s-1}}{z_{s-1}} \end{pmatrix}$$

首先来证明下面关于行列式的引理:

**引理**　命 $a_i>0, 1\leqslant i\leqslant s_i$, 及

$$\Delta_s=\begin{vmatrix} a_1+a_2, & -a_2 & & 0 & \\ -a_2 & a_2+a_3 & \ddots & & \\ & \ddots & \ddots & a_{s-2}+a_{s-1} & -a_{s-1} \\ 0 & & & -a_{s-1} & a_{s-1}+a_s \end{vmatrix}$$

则

$$\Delta_s=a_1\cdots a_s\sum_{i=1}^{s}a_i^{-1}.$$

**证** 当 $s=2$ 时, 命题显然成立. 用归纳法. 假定 $s\geqslant 3$ 及当 $2\leqslant t\leqslant s-1$, 引理的结论对于 $\Delta_t$ 成立, 则

$$\Delta_s=(a_1+a_2)\begin{vmatrix} a_2+a_3, & -a_3 & & 0 & \\ -a_3 & a_3+a_4 & \ddots & & \\ & \ddots & \ddots & a_{s-2}+a_{s-1} & -a_{s-1} \\ 0 & & & -a_{s-1} & a_{s-1}+a_s \end{vmatrix}$$

$$-a_2^2\begin{vmatrix} a_3+a_4, & -a_4 & & 0 & \\ -a_4 & a_4+a_5 & \ddots & & \\ & \ddots & \ddots & a_{s-2}+a_{s-1} & -a_{s-1} \\ 0 & & & -a_{s-1} & a_{s-1}+a_s \end{vmatrix}$$

$$=(a_1+a_2)a_2\cdots a_s\sum_{i=2}^{s}a_i^{-1}-a_2^2a_3\cdots a_s\sum_{i=3}^{s}a_i^{-1}$$

$$=a_1\cdots a_s\sum_{i=2}^{s}a_i^{-1}+a_2^2a_3\cdots a_s\sum_{i=2}^{s}a_i^{-1}-a_2^2a_3\cdots a_s\sum_{i=3}^{s}a_i^{-1}$$

$$=a_1\cdots a_s\sum_{i=1}^{s}a_i^{-1}.$$

引理证完.

$\Delta$ 的计算: 命 $\boldsymbol{A}$ 为 $(s-1)\times(s-1)$ 方阵, 且满足 $\det\boldsymbol{A}=\pm 1$, 则当 $\boldsymbol{H}$ 换为 $\boldsymbol{AH}$ 时, $\Delta$ 是不变的. 命 $\boldsymbol{A}_{ij}=(a_{uv}), 1\leqslant u,v\leqslant s-1$, 此处 $a_{uu}=1$, $1\leqslant u\leqslant s-1$, $a_{ij}=g$, 及 $a_{uv}=0$ (其他情况), 则 $\boldsymbol{H}$ 与 $\boldsymbol{A}_{ij}\boldsymbol{H}$ 的差异为 $\boldsymbol{A}_{ij}\boldsymbol{H}$ 的 $i$- 行等于 $\boldsymbol{H}$ 的 $i$- 行加上 $\boldsymbol{H}$ 的 $j$- 行的 $g$ 倍. 运用这些算子, 即可得到矩阵 $\boldsymbol{A}$ 使

$$\boldsymbol{AH}=\begin{pmatrix} -d_1^{-1} & d_2^{-1} & & & \\ & -d_2^{-1}z_1 & d_3^{-1}z_1 & & 0 \\ & & \ddots & \ddots & \\ 0 & & & -d_{s-1}^{-1}z_1\cdots z_{s-2} & d_s^{-1}z_1\cdots z_{s-2} \end{pmatrix}.$$

所以

$$\Delta^2=\det(\boldsymbol{HH}')=\det(\boldsymbol{AHH}'\boldsymbol{A}')=(z_1^{s-2}\cdots z_{s-2})^2\det\boldsymbol{G},$$

其中

$$G=\begin{pmatrix} d_1^{-2}+d_2^{-2} & -d_2^{-2} & & & \\ -d_2^{-2} & d_2^{-2}+d_3^{-2} & \ddots & & 0 \\ & \ddots & \ddots & & -d_{s-1}^{-2} \\ 0 & & & -d_{s-1}^{-2} & d_{s-1}^{-2}+d_s^{-2} \end{pmatrix}.$$

由引理可知

$$\det \boldsymbol{G} = (d_1 \cdots d_s)^{-2} \sum_{i=1}^{s} d_i^2 = C(\boldsymbol{d})^2,$$

所以

$$\Delta = C(\boldsymbol{d}) z_1^{s-2} \cdots z_{s-2}.$$

## §5. 例

**例 1** 假定 $\boldsymbol{a} = (0.6, 0.15, 0.05)$ 与 $\boldsymbol{b} = (0.8, 0.25, 0.15)$, 则 $a = 0.6+0.15+0.05 = 0.8, b = 1.2$ 及 $\boldsymbol{d} = (1, 0.5, 0.5)$. 由 (10), (11) 与 (19) 式可知

$$N(\boldsymbol{d}) = 1 - \frac{1}{2} - \frac{0.5}{2} = 0.25,$$

及

$$\frac{V(S^*(\boldsymbol{d}))}{V(S(\boldsymbol{d}))} = \frac{0.5}{0.25} = 2.$$

因此可以由 $C^2$ 一个有 $n$ 个点的 NT- 网导出 $S(\boldsymbol{d})$(或 $S(\boldsymbol{a}, \boldsymbol{b})$) 上约有 $\left[\frac{n}{2}\right]$ 个点的 NT- 网. 例如, 由 $C^2$ 上一个 17 个点的佳格子点集出发, 其中有 9 个点落于 $S(\boldsymbol{a}, \boldsymbol{b})$ 之中:

$$\begin{aligned}
&\boldsymbol{x}_1 = (0.765\,7, 0.175\,2, 0.059\,1), \quad &&\boldsymbol{x}_2 = (0.740\,6, 0.176\,2, 0.083\,2),\\
&\boldsymbol{x}_3 = (0.723\,3, 0.161\,3, 0.115\,4), \quad &&\boldsymbol{x}_4 = (0.709\,3, 0.227\,4, 0.063\,3),\\
&\boldsymbol{x}_5 = (0.697\,1, 0.207\,5, 0.095\,4), \quad &&\boldsymbol{x}_6 = (0.686\,2, 0.180\,1, 0.133\,6),\\
&\boldsymbol{x}_7 = (0.667\,2, 0.239\,9, 0.093\,0), \quad &&\boldsymbol{x}_8 = (0.658\,6, 0.204\,1, 0.137\,3),\\
&\boldsymbol{x}_9 = (0.635\,5, 0.232\,2, 0.132\,2).
\end{aligned}$$

**例 2** 假定 $\boldsymbol{a} = (0.3, 0.2, 0.1, 0.05)$ 及 $\boldsymbol{b} = (0.6, 0.5, 0.4, 0.2)$, 则 $a = 0.65$ 与 $\boldsymbol{d} = \left(\frac{6}{7}, \frac{6}{7}, \frac{6}{7}, \frac{3}{7}\right)$. 由 (14) 式可知

$$\begin{aligned}
N(\boldsymbol{d}) &= \int_{\frac{1}{7}}^{1} z_1^2 N\left(\frac{6}{7z_1}, \frac{6}{7z_1}, \frac{3}{7z_1}\right) \mathrm{d}z_1\\
&= \int_{\frac{1}{7}}^{\frac{3}{7}} z_1^2 N\left(\frac{6}{7z_1}, \frac{6}{7z_1}, \frac{3}{7z_1}\right) \mathrm{d}z_1 + \int_{\frac{3}{7}}^{\frac{6}{7}} z_1^2 N\left(\frac{6}{7z_1}, \frac{6}{7z_1}, \frac{3}{7z_1}\right) \mathrm{d}z_1\\
&\quad + \int_{\frac{6}{7}}^{1} z_1^2 N\left(\frac{6}{7z_1}, \frac{6}{7z_1}, \frac{3}{7z_1}\right) \mathrm{d}z_1\\
&= I_1 + I_2 + I_3.
\end{aligned}$$

对于 $I_1$, 有 $\frac{3}{7z_1} \geqslant 1$, 所以 $N\left(\frac{6}{7z_1}, \frac{6}{7z_1}, \frac{3}{7z_1}\right) = \frac{1}{2}$ (见 (13) 式). 因此

$$I_1 = \frac{1}{2}\int_{17}^{\frac{3}{7}} z_1^2 \mathrm{d}z_1 = \frac{26}{6 \times 7^3}.$$

对于 $I_2$, 有 $\frac{6}{7z_1} \geqslant 1$, $\frac{3}{7z_1} \leqslant 1$, 所以由 (18) 式可知

$$I_2 = \int_{\frac{3}{7}}^{\frac{6}{7}} z_1^2 \left(\frac{3}{7z_1} - \frac{1}{2}\left(\frac{3}{7z_1}\right)^2\right) \mathrm{d}z_1 = \frac{27}{7^3}.$$

对于 $I_3$, 有 $\frac{6}{7z_1} \leqslant 1, 1 - \frac{6}{7z_1} < \frac{3}{7z_1}, \frac{6}{7z_1} + \frac{3}{7z_1} \geqslant 1$, 所以由 (19) 式得

$$I_3 = \int_{\frac{6}{7}}^{1} \left(-1 + \frac{15}{7z_1} - \frac{81}{2 \times 7^2 z_1^2}\right) z_1^2 \mathrm{d}z_1 = \frac{88}{6 \times 7^3}.$$

因此

$$N(\boldsymbol{d}) = I_1 + I_2 + I_3 = \frac{276}{6 \times 7^3} = 0.134\ 110\ 787 \cdots$$

于是由 (10)、(11) 式可知

$$\frac{V(S^*(\boldsymbol{d}))}{V(S(\boldsymbol{d}))} = \frac{6^{-1}}{276/6 \times 7^3} = \frac{343}{276} = 1.242\ 753\ 623 \cdots$$

这表明由一个 $C^3$ 上 $\left[\frac{343n}{276}\right]$ 个点的 NT- 网可以诱导出 $S(\boldsymbol{d})$ 上接近于 $n$ 个点的一个 NT- 网.

## 参 考 文 献


[1] Wang Y, Fang, K. T. A note on uniform distribution and experimental design. Kexue Tongbao, 1981, **26**: 485~489.

[2] Weyl, H. Uber die Gleichverteilung der Zahlen mod Eins. Math Ann, 1916, **77**: 313~352.

[3] Hua L, K, Wang, Y. Applications of Number Theory to Numerical Analysis. Heidelberg and Beijing: Springer-Verlag and Science Press, 1981.

[4] Fang, K. T, Wang, Y. Number-Theoretic Methods in Statistics. London: Chapman and Hall, 1994.

[5] Cornell, J, A. Experiments with Mixtures, Design, Models, and the Analysis of Mixture Data. New York: Wiley, 1990.

[6] Wang Y, Fang, K. T. Number-theoretic methods in applied statistics. Chinese Ann Math, Ser B, 1990, **11**: 51~65.

[7] Wang Y, Fang, K. T. Number-theoretic methods in applied statistics(II). Chinese Ann Math, Ser B. 1990, **11**: 384~394.

# 统计模拟中的数论方法

王元 ①* 方开泰 ②

① (中国科学院数学与系统科学研究院数学研究所, 北京 100190)
② (中国科学院数学与系统科学研究院应用数学研究所, 北京 100190)

## 摘 要

受现实生活的一个案例形成的随机覆盖问题之启发, 我们在本文中讨论并比较了统计模拟中的几种方法, 包括 ELP 网、NT 网与其他网. 我们给出的一些结果对统计模拟是有用的.

**关键词 数论方法 统计模型 几何概率**

**MSC(2000) 主题分类 65C05, 65C50**

## 1. 引 言

由于统计中的许多问题都没有解析解, 所以统计模拟是一个重要工具. 命 $\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$ 为一个具有 c.d.f $F(x)$ 的总体的一个样本, $T$ 为 $\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$ 的一个统计量. 统计模拟可以产生一个样本并构造出 $T$ 的一个样本 $T_1$. 重复上述过程 $m$ 次, 我们得到 $T$ 的一个样本集 $T_1,\cdots,T_m$. 当 $m$ 大时, $T_1,\cdots,T_m$ 的经验分布渐进于 $T$ 的分布.

作为数论方法中 NT 网在统计模拟中的一个案例, 方开泰与王元 [1,2] 建议了一个固定圆被 $m$ 个随机圆覆盖面积的分布模型, 现在我们将这个问题叙述如下:

命 $B_2=\{(x,y):x^2+y^2\leqslant 1\}$ 为单位圆. 假定有 $m$ 个随机圆 $O_1,\cdots,O_m$, 它们的中心与半径分别为 $P_1,\cdots,P_m$ 与 $R_1,\cdots,R_m$, $P_i$ 相互独立, 且

$$P_i\sim N_2(\boldsymbol{0},\sigma_i^2\boldsymbol{I}_2),\quad 1\leqslant i\leqslant m,$$

①引用格式: 王元, 方开泰. 统计模拟中的数论方法. 中国科学 A, 2009, 39(7): 775—782

Wang Y, Fang K T. On number-theoretic method in statistics simulation. Sci China Ser A, 2009, 52, DOI: 10.1007/s11425-009-0126-3

此处 $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 为具有均值矢量 $\boldsymbol{\mu}$ 与协方差矩阵 $\boldsymbol{\Sigma}$ 的二元正态分布, $\sigma_i > 0, \mathbf{0} = (0, 0), \boldsymbol{I}_2$ 为 $2 \times 2$ 单位矩阵. 命 $S$ 为 $B_2$ 与所有随机圆的并的公共区域, 即

$$S = B_2 \cap (O_1 \cup \cdots \cup O_m).$$

我们亦用 $S$ 表示 $S$ 的面积, 希望求出 $S$ 的分布图. 图 1 表示 $m = 3$ 的一种情况, 其中 $S$ 是有阴影的区域.

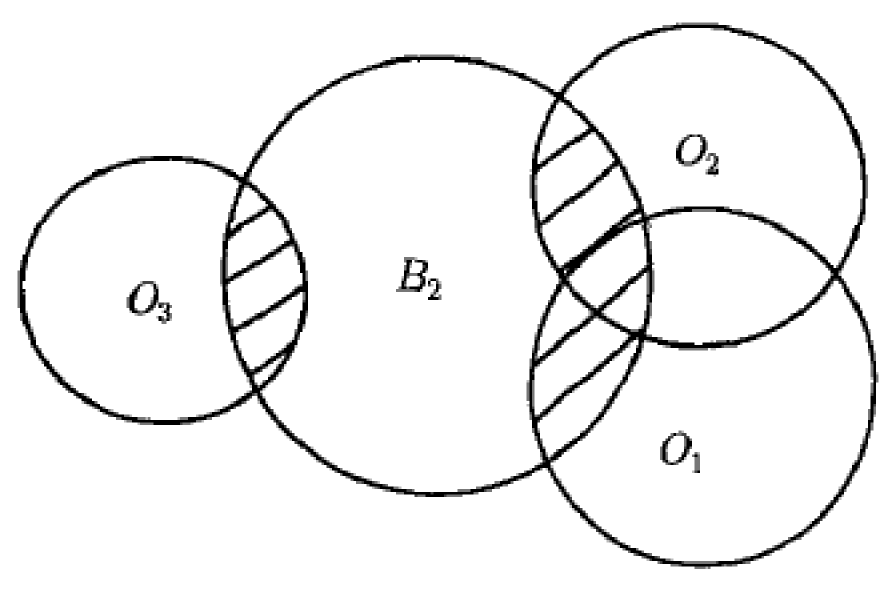



图 1

若 $m = 1$, 由于两个圆的分共面积可以由这两个圆的中心与半径的显式表示出来, 所以容易找到 $S$ 的分布. 当 $m > 1$, 则难于找到 $S$ 分布的一个简单公式.

若 $m = 1$, 我们可以对 $S$ 的模拟结果与 $S$ 的真值来进行比较. 在此, 经典方法, 即等距格点方法 (ELP 网) 与数论方法 (NT 网) 被用来作统计模拟. 一些数值实例表明 NT 网远比 ELP 网为优 (例如见文献 [1, 2]). 为简单计, 我们可以取 $m = 0$, 即 $S$ 为单位圆的面积 $\pi$, 我们仅对模拟结果与 $\pi$ 进行比较即可 (见文献 [3, 例 3a]).

本文的目的在于用几何数论的结果来对统计模拟中的几种方法作些比较说明.

## 2. ELP 网与 NT 网

命 $ABCD$ 为 $S_2$ 的外接正方形, 如图 2 所示. 将 $ABCD$ 分割成边长为 $2/n$ 的 $n^2$ 个面积相等的正方形, 则我们得到 $ABCD$ 中 $n^2$ 个格点

$$\left\{ \left( -1 + \frac{2i}{n}, -1 + \frac{2j}{n} \right), 0 \leqslant i, j \leqslant n-1 \right\}. \tag{1}$$

点集 (1) 称为一个 ELP 网, 假定共有 $N$ 个格子点落于 $B_2$ 中, 现在我们用标准 Monte Carlo 方法生成 $m$ 个随机圆, 它们的圆心与半径分别为 $P_i$ 与 $R_i$ $(1 \leqslant i \leqslant m)$. 假定在这 $N$ 个点中有 $M$ 个点落于区域 $O_1 \cup \cdots \cup O_m$ 之中, 则我们得到 $S$ 的一个近似的观测值 $\pi M/N$. 重复这一步骤, 我们得到 $S$ 的一个近似的观测值系, 从而获得 $S$ 的一个经验分布. 我们称这个模拟程序为方法 I.

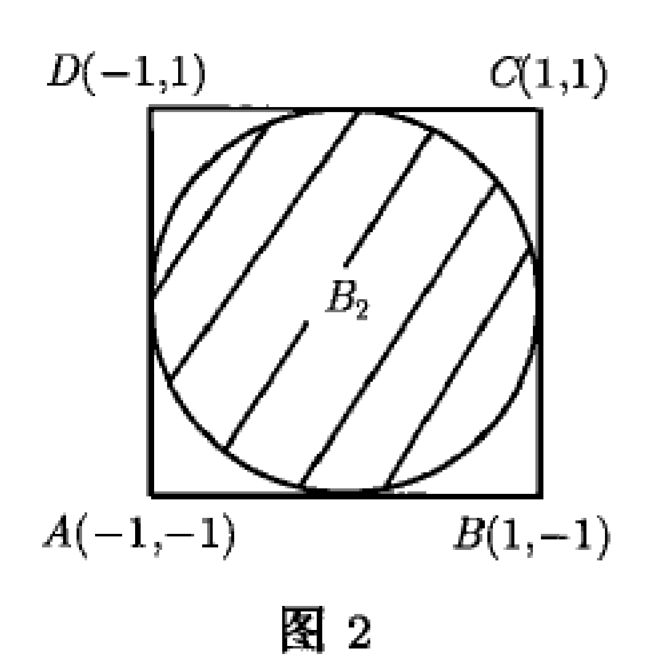



图 2

若用一个 NT 网代替集合 (1). 例如取 $ABCD$ 上的一个佳格点网 (好格子点网, glp 网), 并如前所述进行模拟, 则称这个程序为方法 II.

这两个方法都基于 $ABCD$ 上的均匀散布点集, 而不是直接用 $B_2$ 上的均匀散布点集. 我们当然可以用 $B_2$ 上的一个网来进行模拟. 变换

$$\begin{cases} x = r\cos 2\pi\theta, \\ y = r\sin 2\pi\theta, \end{cases}$$

将单位正方形 $U_2 = \{(r,\theta): 0 \leqslant r,\theta \leqslant 1\}$ 映射为 $B_2$, 我们有

$$\int\int_{x^2+y^2\leqslant 1} dxdy = \int\int_{U^2} 2\pi r dr d\theta. \tag{2}$$

命 $P_q = \{P_q(i) = (r_i,\theta_i), 1 \leqslant i \leqslant q\}$ 为 $U_2$ 上的一个均匀散布点集 (NT 网). 则

$$Q_q = \{Q_q(i) = (r_i\cos 2\pi\theta_i, r_i\sin 2\pi\theta_i), 1 \leqslant i \leqslant q\}$$

为 $B_2$ 上的 $q$ 个点所成之点集. 注意 $\{Q_q\}$ 在 $B_2$ 上不是均匀散布的. 但是由 (2) 可知我们可以用加权和 $\sum_{i=1}^{q} 2\pi r_i = N^*$ 来代替 $N$. 假定 $Q_q(i_j)(1 \leqslant j \leqslant t)$ 为 $\{Q_q(i)\}$ 中被 $O_1 \cup \cdots \cup O_m$ 覆盖的诸点, 则 $\sum_{j=1}^{t} 2\pi r_{i_j} = M^*$ 可以用来代替 $M$. 因此我们可以如前一样来进行模拟, 并称这个程序为方法 III.

我们可以直接定义 $B_2$ 上的 NT 网如下: 命 $P_q$ 为 $U_2$ 上的一个 NT 网, 它有 Weyl[4] 意义下的偏差

$$D(q) = \sup_{(r,\theta)\in U_2} \left| \frac{N(r,\theta)}{q} - r\theta \right|, \tag{3}$$

此处 $N(r,\theta)$ 表示满足 $r_i \leqslant r, \theta_i \leqslant \theta$ 的点 $P_q(i) = \{(r_i,\theta_i), 1 \leqslant i \leqslant q\}$ 的个数. 则 $\{(x_i,y_i), 1 \leqslant i \leqslant q\}$ 是一个 $B_2$ 上的均匀散布点集, 其中

$$\begin{cases} x_i = \sqrt{r_i}\cos 2\pi\theta_i, \\ y_i = \sqrt{r_i}\sin 2\pi\theta_i, \quad 1 \leqslant i \leqslant q. \end{cases}$$

我们可以证明

$$\sup_R \left| \frac{N(R)}{q} - \theta r \right| = D(q), \tag{5}$$

此处 $N(R)$ 表示落入扇形区域 $\{r_i \leqslant r, \theta_i \leqslant \theta\}$ 中的点数, 如图 3 所示 [1,2]. (5) 式的左端称为 $B_2$ 上的集合 (4) 的 $F$- 偏差, 它是用来度量 $B_2$ 上的集合均匀性的测度. (5) 式表示 $B_2$ 上的集合 (4) 的 $F$- 偏差, 它等于 $U_2$ 上的集合 $P_q(i)$ 的偏差. 因此我们可以直接用集合 (4) 来做统计模拟, 这个程序被称为方法 IV.

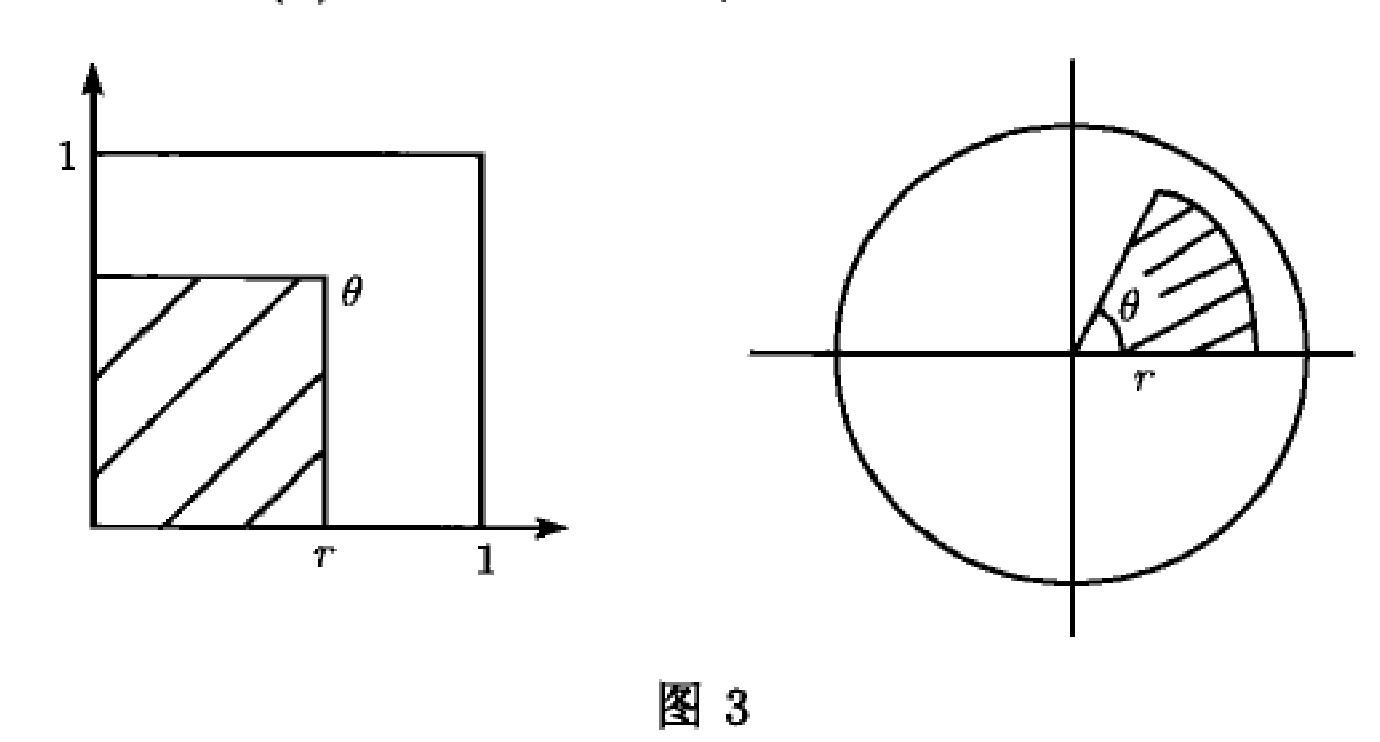



图 3

用 ELP 网与 NT 网 (glp 网) 做模拟的一些数值算例表明方法 II, III 与 IV 皆比方法 I 为优, 而方法 III 与 IV 具有同等精密度, 且均比方法 II 为优 [1,2].

## 3. 圆 问 题

命 $A_2(x)$ 表示圆 $\{(u,v); u^2+v^2 \leqslant x\}$ 内格点或整点 $(u,v)$ 的个数. Gauss[5] 证明了

$$A_2(x) - \pi x = O(x^{1/2}). \tag{6}$$

命 $\theta$ 为使关系式

$$A_2(x) - \pi x = O(x^v)$$

成立的 $v$ 的下界. 寻求 $\theta$ 的最佳估计是一个几何数论的著名问题, 称为圆问题. 估计 (6) 表示 $\theta \leqslant 1/2$. 这一结果不断地被改进, 例如 Sierpinski[6] ($\theta \leqslant 1/3$), Van der Corput[7] ($\theta \leqslant 37/112$), Titchmarsh[8] ($\theta \leqslant 15/46$), 华罗庚 [9] ($\theta \leqslant 13/40$), 陈景润 [10] ($\theta \leqslant 12/37$) 等. 目前, 最佳记录是由 Iwaniec 与 Mozzochi[11] 得到的:

$$A_2(x) - \pi x = O(x^{\frac{7}{22}+\varepsilon}), \quad \varepsilon > 0, \tag{7}$$

这表示 $\theta \leqslant 7/22$. 另一方面, Hardy[12] 证明了 $\theta \geqslant 1/4$, 更精确地说, 他证明了

$$\overline{\lim_{x\to\infty}} \frac{A_2(x) - \pi x}{x^{\frac{1}{4}} \log^{\frac{1}{4}} x} > 0, \tag{8}$$

将 (6) 式的两端均除以 $x$, 则得

$$\frac{A_2(x)}{x}-\pi=O(x^{-1/2}), \tag{9}$$

此处 $A_2(x)/x$ 可以看作 $B_2$ 中以 $1/\sqrt{x}$ 为边长的所有方形的面积之和. 记 $N_2=A_2(x)/x$. 则 (9) 可以写成

$$N_2-\pi=O(x^{-\frac{1}{2}}). \tag{10}$$

但由 (8) 可知 (10) 的右端不能改进得比

$$O(x^{\frac{-3}{4}}\log^{\frac{1}{4}}x) \tag{11}$$

更好. 这是方法 I(ELP 网) 精密度的极限.

若在 (5) 中取 $\theta=1$, 则扇形区域 $R$ 为以原点为中心、$r$ 为半径的圆, 如图 4 所示. 命 $N(r)$ 为点集 (4) 落入圆 $\{(u,v):u^2+v^2\leqslant r^2\}$ 中的点集, 则由 (5) 可知

$$\sup_r\left|\frac{N(r)}{q}\pi r^2-\pi r^2\right|=O(D(q)). \tag{12}$$

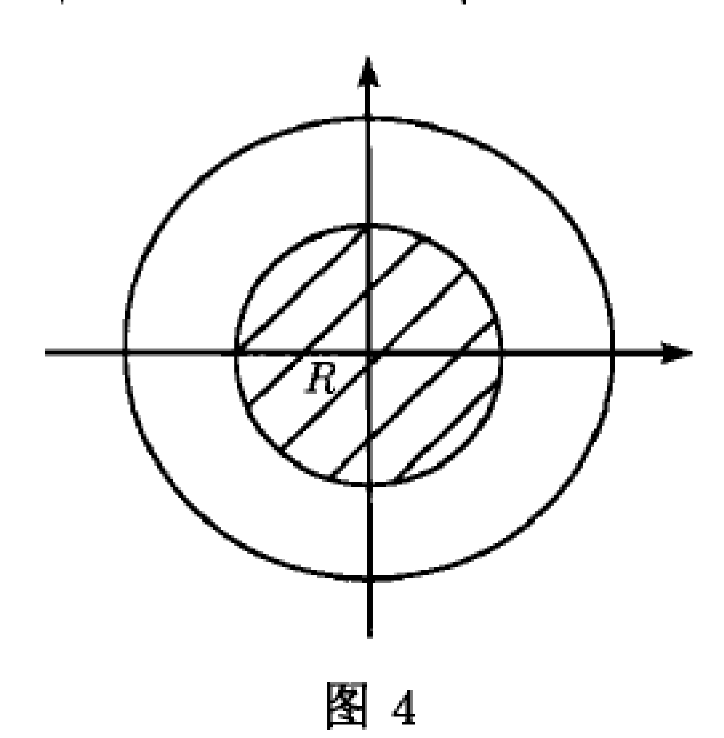



图 4

我们引入一个有低偏差 $D(q)$ 的 $U_2$ 上的一个 NT 网, 及由此诱导出一个 $B_2$ 上的 NT 网, 具有 $F$- 偏差 $D(q)$, 如下:

命 $\{F_n\}$ 为 Fibonacci 序列, 此处诸 $F_n$ 为由下面的递推公式定义的整数

$$F_0=0,F_1=1,F_{n+1}=F_n+F_{n-1}\quad(n\geqslant 1),$$

或直接由关系式

$$F_n=\frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n-\left(\frac{1-\sqrt{5}}{2}\right)^n\right),\quad n=0,1,\cdots$$

定义. Bahvalov[13], 华罗庚与王元 [14] 独立地建议了 $U_2$ 上的点集

$$\left(\frac{k}{F_n},\left\{\frac{kF_{n-1}}{F_n}\right\}\right),\quad 1\leqslant k\leqslant F_n, \tag{13}$$

其中 $n \geqslant 3$, $\{y\}$ 表示 $y$ 的分数部分. 他们还用集合 (13) 来构造 $U_2$ 上的求积公式. Zaremba[15] 证明了集合 (13) 有偏差

$$D(F_n) = O\left(\frac{\log F_n}{F_n}\right). \tag{14}$$

由 Schmidt[16] 关于均匀分布的著名定理可知 (14) 的右端是臻于至善的. 如果我们用 (13) 诱导的 $B_2$ 上的网作模拟, 则仅导致误差 $O(\log F_n/F_n)$. 这说明 ELP 网与 NT 网给出的模拟误差分别为

$$O(q^{-\frac{3}{4}}\log^{\frac{1}{4}} q) \quad 与 \quad O(q^{-1}\log q), \tag{15}$$

此处 $q$ 为网中所含点数, 这可以看作为什么在模拟中用 NT 网比 ELP 网为佳这个问题的说明.

**注记 1** ELP 网的偏差不会比 $O(q^{-1/2})$ 更好 [17]. 因此若一个 ELP 网映射至 $B_2$, 则诱导的集合仅有 $F$- 偏差 $O(q^{-1/2})$.

## 4 高维球问题

第一节引进的统计模拟模型可以推广至 $s(>2)$ 维情形. 命 $B_s$ 为 $s$ 维单位球:

$$B_s = \{\boldsymbol{x} = (x_1, \cdots, x_s) : x_1^2 + \cdots + x_s^2 \leqslant 1\}.$$

假定有 $m$ 个随机球 $O_1, \cdots, O_m$, 它们的中心与半径分别为 $P_1, \cdots, P_m$ 与 $R_1, \cdots, R_m$, $P_1$ 相互独立, 且 $P_i \sim N_s(\mathbf{0}, \sigma_i^2 \boldsymbol{I}_s)$, 其中 $N_s$ 为多元正态分布, $\sigma_i > 0, \mathbf{0} = (0, \cdots, 0)'$ 及 $\boldsymbol{I}_s$ 为 $s \times s$ 单位矩阵. 试求 $S = B_s \cap (O_1 \cup \cdots \cup O_m)$ 的体积的分布, 此处我们仍用 $S$ 表示 $S$ 的体积.

我们用 $A_s(x)$ 表示球 $\{(x_1, \cdots, x_s) : x_1^2 + \cdots + x_s^2 \leqslant x\}$ 中的整点 $(x_1, \cdots, x_s)$ 个数. 则类似于 Gauss 圆问题, 我们可以证明

$$A_s = v(B_s)x^{s/2} + O(W), \tag{16}$$

此处

$$v(B_s) = \frac{\pi^{s/2}}{\Gamma(\frac{s}{2}) + 1} \tag{17}$$

为 $B_s$ 的体积, $W = O(x^{\varphi})$, 其中 $\varphi$ 为一个满足 $0 < \varphi < s/2$ 的数.

当 $s = 3$ 时, 陈景润 [18] 与 Vinogradov[19] 独立地证明了 $\varphi = \dfrac{2}{3} + \varepsilon(\varepsilon > 0)$. 但是我们有 $W = \Omega(x^{1/2}\log^2 x)$, 这就是说, 应用 ELP 网的误差极限为

$$N_3 - v(B_3) = O(q^{-2/3}\log^2 q), \tag{18}$$

此处 $N_3 = A_3(x)/q, q = [x^{3/2}]$, 其中 $[y]$ 表示 $y$ 的整数部分.

当 $s \geqslant 4$ 时, Walfisz[20] 与 Landau[21] 证明了 (16) 式的误差项适合

$$W = \begin{cases} O(x\log^2 x), & 当\ s = 4, \\ O(x^{s/2-1}), & 当\ s > 4. \end{cases} \tag{19}$$

Jarnik[22] 证明了

$$W = \Omega(x^{s/2-1}), \quad s \geqslant 4.$$

因此除去 $s = 4$ 的对数阶外, 估计 (19) 是臻于至善的.

将 (16) 的两端除以 $x^{s/2}$ 则得

$$N_s - v(B_s) = O(q^{-2/s}), \tag{20}$$

此处 $N_s = A_s(x)/x^{s/2}, q = [x^{s/2}]$, 其中在计算 $s = 4$ 的情况时, 对数项被忽略了, 这就是 ELP 网的极限.

现在我们取一个 $U_s$ 上的 NT 网, 并推荐使用 Korobov[23] 与 Hlawka[24] 独立建议的 glp 网:

$$\left(\left\{\frac{a_1 k}{q}\right\}, \cdots, \left\{\frac{a_s k}{q}\right\}\right), \quad 1 \leqslant k \leqslant q, \tag{21}$$

此处 $\mathbf{a} = \{a_1, \cdots, a_s\}$ 为一个整矢量. 他们证明了, 当 $q = p$ 为素数时, 存在一个矢量 $\mathbf{a}$ 使集合 (21) 有偏差

$$D(p) = O\left(\frac{\log^s p}{p}\right). \tag{22}$$

我们将集合 (21) 记为 $\{\boldsymbol{c}_k = (c_{k1}, \cdots, c_{ks}), 1 \leqslant k \leqslant q\}$. 当 $s = 3$ 时, 我们定义

$$\begin{cases} x_{k1} = c_{k1}^{1/3}(1 - 2c_{k2}), \\ x_{k2} = 2c_{k1}^{1/2}\sqrt{c_{k2}(1 - c_{k2})}\cos(2\pi c_{k3}), \\ x_{k2} = 2c_{k1}^{1/3}\sqrt{c_{k2}(1 - c_{k2})}\sin(2\pi c_{ks}), \quad 1 \leqslant k \leqslant q. \end{cases}$$

若 $q = p$ 为一个素数, 则存在 $\boldsymbol{a} = (a_1, a_2, a_3)$ 使 $\{\boldsymbol{x}_k = (x_{k1}, x_{k2}, x_{k3}), 1 \leqslant k \leqslant p\}$ 为 $B_s$ 上的一个 NT 网, 其中 $F$- 偏差为 $O(\log^3 p/p)$.

当 $s > 3$ 时, 我们取 $U_3$ 上含有 $p$ 个元素的一个 glp 网. 记它为 $\{\boldsymbol{c}_k = \{(c_{k1}, \cdots, c_{ks}), 1 \leqslant k \leqslant p\}$. 命

$$F_j(\varphi) = \begin{cases} \varphi^s, & 当\ j = 1, \\ \dfrac{\pi}{B(\frac{1}{2}, \frac{s-j+1}{2})}\displaystyle\int_0^{\varphi}(\sin \pi t)^{s-j}dt, & 当\ 2 \leqslant j \leqslant s, \end{cases}$$

此处 $B(a,b)$ 为 Beta 函数. 记

$$\begin{cases} b_{k1}=c_{k1}^{1/s}, \\ b_{ki}=F_i^{-1}(c_{ki}), \quad 2\leqslant i\leqslant s, \quad 1\leqslant k\leqslant p, \end{cases}$$

此处 $F_i^{-1}(x)$ 为 $x_i$ 的 c.d.f. $F_i(x)$ 的逆函数. 定义

$$\begin{cases} x_{kj}=b_{k1}\prod_{i=2}^{j} S_{ki}C_{k,j+1}, \quad 1\leqslant j\leqslant s-1, \\ x_{ks}=b_{k1}\prod_{i=2}^{s} S_{ki}, \end{cases}$$

此处

$$S_{ki}=\sin(\pi b_{ki}), \quad C_{ki}=\cos(\pi b_{ki}), \quad 2\leqslant i\leqslant s-1,$$

$$S_{ks}=\sin(2\pi b_{ks}), \quad C_{ks}=\cos(2\pi b_{ks}), \quad 1\leqslant k\leqslant p.$$

则 $\boldsymbol{x}_k=(x_{k1},\cdots,x_{ks}), 1\leqslant k\leqslant p$ 为 $B_s$ 上的一个 NT 网, 它的 $F$- 偏差等于 $\boldsymbol{c}_k$ 的偏差, 即有阶 [1,2]

$$O(p^{-1}\log^s p). \tag{23}$$

(22) 与 (23) 是

$$O(q^{-2/s}) \quad 与 \quad O(q^{-1}\log^s q) \tag{24}$$

的比较, 这可以看作在统计模拟中, NT 网比 ELP 网优越的说明.

**注记 2** 由均匀分布的一个重要猜想可知, 对于 $U_s$ 上任何 $q$ 个点的集合, 其偏差皆适合 $D(q)=\Omega(q^{-1}\log^{s-1}q)$.

**注记 3** Halton 网可以达到估计 $D(q)=O(q^{-1}\log^{s-1}q)$, 但是 Halton 网的构造比 glp 网要复杂 [25].

**注记 4** 对于一个素数 $p$, glp 网的构造依赖于矢量 $\boldsymbol{a}=\boldsymbol{a}(p)$. 但求出 $\boldsymbol{a}(p)$ 需要 $O(p^2)$ 次初等运算, 所以由数值分析的角度看, 这个方法只是存在性结果. 我们将 Fibonacci 网 (13) 推广至 $U_s(s>2)$, 但其偏差却大于 (23)[17].

## 5. 其 他 区 域

首先, 对于有理椭球

$$Q(\boldsymbol{x})=\sum_{i=1}^{s}\sum_{j=1}^{s}a_{ij}x_ix_j,$$

此处 $a_{ij}=a_{ji}$ 及 $a_{ij}\in Q$, 存在一个有理系数的线性变换 $T$, 它将 $Q(\boldsymbol{x})$ 映射至 $B_s$. 因此第四节所述的结果对 $Q(\boldsymbol{x})$ 亦成立 (见文献 [26]). 当 $s=2$ 时, Jarnik[22] 证明了下面定理:

命 $L$ 为一个闭 Jordan 曲线, 它的长度亦记为 $L$. 假定 $L \geqslant 1$, 命 $A$ 为由 $L$ 包围的区域, 此处 $A$ 的面积亦记为 $A$. 命 $N$ 为 $A$ 中整点之个数, 则

$$|A - N| < L. \tag{25}$$

这个结果的精密度等于 Gauss 关于圆的问题的精密度 (见 (6)). 取 $A$ 为单位正方形, 则由 $A$ 上 ELP 网的偏差之上, 下界估计可知 (25) 右端的无穷大阶是臻于至善的.

若 $s > 2$, 则由 $U_s$ 上 ELP 网的偏差可知 (25) 的右端应不会优于 $A$ 的边界的 $(s-1)$ 维测度.

命 $A$ 为一个 $s$ 维有界区域. 假定存在一个变换将 $U_s$ 映射至 $A$, 则由逆变换法, 我们可得到一个由 $U_s$ 上的 NT 网 $\tilde{U}_s$ 诱导出来的 $A$ 上的一个 NT 网 $\tilde{A}$, 其中 $\tilde{A}$ 的 $F$- 偏差等于 Weyl 意义之下 $\tilde{U}_s$ 的偏差 (见文献 [1,2]). 因此我们建议用 $\tilde{A}$ 来作 $A$ 上统计量的统计模拟, 即我们推荐用第二节所讲的方法 IV.

若 $A$ 是 $s$ 维空间的一个有界区域, 但其维数 $t$ 小于 $s$, 例如 $A$ 是 $B_s$ 的边界 $S_{s-1}$, 则我们需要一个将 $U_t$ 映射至 $A$ 的映射 $\varphi$. 现在我们可以用 $A$ 的体积元素去定义 c.d.f. $F(\boldsymbol{x})$, 然后我们可以决定由 $U_t$ 上的一个 NT 网 $\tilde{U}_t$ 诱导出来的 $A$ 上的 NT 网 $\tilde{A}$, 从而仍可以如前一样应用方法 IV[1,2].

我们有由 $U_t$ 至下面每一个区域的映射:

$A_s = \{\boldsymbol{x} = (x_1, \cdots, x_s) : 0 \leqslant x_1 \leqslant \cdots \leqslant x_s \leqslant 1\}$,

$B_s = \{\boldsymbol{x} = (x_1, \cdots, x_s) : x_1^2 + \cdots + x_s^2 \leqslant 1\}$,

$S_{s-1} = \{\boldsymbol{x} = (x_1, \cdots, x_s) : x_1^2 + \cdots + x_s^2 = 1\}$,

$V_s = \{\boldsymbol{x} = (x_1, \cdots, x_s) \in \mathbb{R}_s^+ : x_1 + \cdots + x_s \leqslant 1\}$,

$T_{s-1} = \{\boldsymbol{x} = (x_1, \cdots, x_s) \in \mathbb{R}_s^+ : x_1 + \cdots + x_s = 1\}$,

$T_{s-1}(\boldsymbol{a}, \boldsymbol{b}) = \{\boldsymbol{x} = (x_1, \cdots, x_s) \in \mathbb{R}_s^+ : x_1 + \cdots + x_s = 1, 0 \leqslant a_i \leqslant x_i \leqslant b_i \leqslant 1, 1 \leqslant i \leqslant s\}$,

此处 $R_s^+$ 表示所有非负元素矢量构成的集合, 及 $a_i's$ 与 $b_i's$ 为 [0, 1] 中的常数 (见文献 [1, 2, 27]).

若 $A$ 是一个 $s$ 维区域, 但我们不能明确写出由 $U_s$ 至 $A$ 的映射, 则我们可以定义一个包围 $A$ 的长方体 $R$. 然后利用 $R$ 上的一个 NT 网来对 $A$ 上的统计量作统计模拟, 即用第二节引进的方法 II. 一般说来, 这个方法的误差大于方法 IV 的误差, 但仍比 GLP 网为优 (即方法 I).

**致谢** 感谢周永道博士和宋谢冰小姐协助中文 Latex 输入工作, 感谢周宏先生的许多帮助.

## 参 考 文 献


[1] Fang K T, Wang Y. Number-Theoretic Methods in Statistics. Chapman and Hall, 1994

[2] Wang Y, Fang K T. Number theoretic methods in applied statistics. *Chin Ann Math Ser B*, **11**: 41–55 (1990); II, *Chin Ann Math Set B*, **11**: 384–394(1990)

[3] Ross S D. A Course in Simulation. New York: Macmillan Publishing Company, 1991

[4] Weyl H. Über die gleichverteilung der zahlen mod Eins. *Math Ann*, **77**: 313–352 (1916)

[5] Gauss C F. De nexu inter multitudinem classium etc. *Werke*, **2**: 269–291(1863)

[6] Sierpinski W. O pewnem zagadniense z rachunka funkcyi asymptotycznych. *Prace Mat-Fiz*, **17**: 77–118(1906)

[7] van der Corput J G. Neue zahlentheoretische abschatzungen. I, *Math Ann*, **89**: 215–254(1923); II, *Math Z*, **29**: 397–428(1928)

[8] Titchmarsh E C. The lattice points in a circle. *Proc London Math Soc*, **38**: 96–115(1935)

[9] Hua L K. The lattice points in a circle. *Quart J Math Oxford*, **13**: 18–29(1942)

[10] Chen J R. The circle problem. *Acta Math Sin*, **13**: 299–313(1963)

[11] Iwaniec H, Mozzochi C J. On divisor and circle problems. *J Numer Theory*, **29**: 60–93(1988)

[12] Hardy G H. On Dirichlet's divisor prolem. *Proc London Math Soc*, **15**: 1–25(1916)

[13] Bahvalov N S. Approximate computation of multiple integrals. *Ves Mos Univ Set Math*, **4**: 3–18(1959)

[14] Hua L K, Wang Y. Remarks concerning numerical integration. *Sci Rec*, **4**: 8–11(1960)

[15] Zaremba S K. Good lattice points, discrepancy and numerical integration. *Ann Mat Pura Appl* (4), **73**: 293–318(1966)

[16] Schmidt W M. Irregularities of distribution, VII. *Acta Arith*, **21**: 45–50(1972)

[17] Hua L K, Wang Y. Applications of Number Theory to Numberical Analysis. Berlin/ Beijing: Springer-Verlag/Science Press, 1981

[18] Chen J R. Improvement of asymptotic formulas for the number of lattic points in a region of the three dimensions(II). *Sci Sin*, **12**: 751-764(1963)

[19] Vinogradov I M. The number of integral points in a sphere. *Izv AN SSSR*, **27**: 957–968(1963)

[20] Walfisz A. Über Gitterpunkte in mehrdimensionalen ellipsoiden. *Math Z*, **19**: 300–307 (1924)

[21] Landau E. Über gitterpunkte in mehrdimensionalen ellipsoiden. *Math Z*, **21**: 126-132(1924)

[22] Jarnik V. Über Gitterpunkte in der Ebene. *Rospravy*, **33**: (1924)

[23] Korobov N M. The approximate computation of multiple integrals. *DAN SSSR*, **124**: 1027-1210(1959)

[24] Hlawka E. Zur angenaherten Berechnung mehrfacher Integrale. *Mon Math*, 66: 140–151(1962)

[25] Halton J H. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer Math*, **2**: 84–90(1960)

[26] Hua L K. On Estimation of Exponential Sums and Their Applications in Number Theory. Enz Teubner Verlag, 1959

[27] Wang Y, Fang K T. Uniform design of experiments with mixtures. *Sci Sin*, **26**: 1–10(1996)

(O-3874.0101)

华罗庚文集 | 应用数学卷 I |


ISBN 978-7-03-027251-5

9 787030 272515 >

销售分类建议：高等数学

定 价：98.00元